Cooperative Estimation of 3D Target Motion via Networked Visual Motion Observer

Takeshi Hatanaka, *Member, IEEE* and Masayuki Fujita, *Member, IEEE*

Abstract

This paper investigates cooperative estimation of 3D target object motion for visual sensor networks. In particular, we consider the situation where multiple smart vision cameras see a group of target objects. The objective here is to meet two requirements simultaneously: averaging for static objects and tracking to moving target objects. For this purpose, we present a cooperative estimation mechanism called networked visual motion observer. We then derive an upper bound of the ultimate error between the actual average and the estimates produced by the present networked estimation mechanism. Moreover, we also analyze the tracking performance of the estimates to moving target objects. Finally the effectiveness of the networked visual motion observer is demonstrated through simulation.

Index Terms

Cooperative estimation, Visual-based observer, Averaging, Passivity, Visual sensor network

I. INTRODUCTION

A visual sensor network [1], [2] is a kind of wireless sensor network consisting of spatially distributed smart cameras with communication and computation capability. Unlike other sensors measuring values such as temperature and pressure, vision sensors do not provide explicit data but combining image processing techniques or human operators gives rich information on situation awareness such as what happens, what a target is, where it is and where it bears. Due to their nature, visual sensor networks are useful in environmental monitoring, surveillance, target tracking and entertainment and are expected as a component of sustainable infrastructures.

Takeshi Hatanaka(corresponding author) and Masayuki Fujita are with the Department of Mechanical and Control Engineering, Tokyo Institute of Technology, Tokyo 152-8550, JAPAN, hatanaka@ctrl.titech.ac.jp, fujita@ctrl.titech.ac.jp

A lot of research works have been devoted to fusing control techniques with visual information so-called visual feedback control or images in the loop [3]–[9]. The motivating scenarios of the fusion currently spread over the robotic systems into security and surveillance systems, medical imaging procedures, human-in-the-loop systems and even understanding biological perceptual information processing. Driven by the technological innovations of the smart wearable cameras, the aforementioned networked vision system also emerges as a challenging new application field of the visual feedback control and estimation.

In this paper, we focus on estimation of 3D rigid body motion as in [7]–[9], and reconsider the problem not for a single camera system but for the networked vision systems. In particular, we aim at an extension of [8] from the single camera to visual sensor networks, where the paper [8] presents a vision-based observer called *visual motion observer* [9] estimating 3D target object motion from 2D vision data. In visual sensor networks, it is expected that not only an estimate is produced but also the vision cameras cooperate with each other in an efficient manner, which brings us new theoretical challenges. The advantages of cooperation are: (i) accurate estimation by integrating rich information, (ii) tolerance against obstruction, misdetection in image processing and sensor failures and (iii) wide vision and elimination of blind areas by fusing images of a scene from a variety of viewpoints. To tackle such distributed estimation problems, cooperative control as in [10]–[15] provides useful methodologies. In this paper, we especially focus on passivity-based cooperative control schemes investigated in [12]–[15].

Cooperative estimation for sensor networks has been addressed in [16]–[24]. The main objective of these researches is averaging the local measurements or local estimates among sensors in a distributed fashion in order to improve estimation accuracy. For this purpose, most of the works utilize the consensus protocol [10] in the update of the local estimates. While [16], [17] assume that parameters to be estimated are fixed, [18]–[24] address estimation of dynamic parameters assuming that the parameters follow some dynamical system. Among them, [18]–[22] execute a large number of consensus iterations between each update of estimates, which is hardly applicable to dynamic estimation problems except for the case of slow dynamics. Meanwhile, [23] and [24] present estimation algorithms without using such iterations. Unfortunately, however, most of these algorithms are not applicable to our problem since the object's pose takes values in a non-Euclidean space and the consensus scheme on a vector space [10] does not work there.

Meanwhile, average computation in the group of rotations is tackled by [17], [25], [26].

The paper [25] defines two types average rotations, Euclidean and Riemannian means, and derives their fundamental properties. Reference [26] presents a computational algorithm of the Riemannian mean and analyzes its convergence. The paper [17] presents a distributed version of the algorithm in [26] based on the consensus protocol [10], which is motivated by the visual sensor networks. However, [17] focuses on averaging by assuming that the target orientations are obtained *a priori* and the scheme cannot be essentially extended to dynamic estimation problems.

In this paper, we present a novel cooperative estimation mechanism called *networked visual motion observer*. We consider the situation where multiple smart vision cameras capture a group of target objects. Under the situation, the objective of the present estimation mechanism is to meet two requirements simultaneously: *averaging for static objects*, which means gaining estimates close to an average of multiple target objects' poses, and *tracking to moving target objects*, which means that the estimates track the moving average within a bounded error. Namely, the present mechanism deals with both static and dynamic estimation problems. For this purpose, we first present the *networked visual motion observer*, which consists of the visual feedback and mutual feedback from neighboring vision cameras, based on the passivity-based visual motion observer [8] and the passivity-based pose synchronization law presented in [15].

We next evaluate the averaging performance attained by the networked visual motion observer. For this purpose, we define a notion of approximate averaging by using the ultimate error between the actual average and the estimates produced by the present observer. Then, we derive an upper bound of the ultimate error, whose partial solution is already given in [27], [28] and this paper provides its generalized version. The result gives us an insight into the gain selection such that average estimation becomes accurate if mutual feedback is much stronger than visual feedback.

We moreover evaluate the tracking performance of the estimates to moving target objects. Here, we view the body velocities of the target objects as a disturbance of the total networked system and evaluate the ultimate distance from the estimates to the average. We see from the result an insight that choosing a large visual feedback gain results in a good tracking performance.

Finally, we demonstrate the effectiveness of the present networked visual motion observer and validity of the theoretical results through simulation.

The organization of this paper is as follows. Section II explains the situation under consideration in this paper and formulates the visual sensor networks together with the objective to be met. In Section III, after introducing the visual motion observer [8], we present the networked

visual motion observer. Section IV clarifies accuracy of the average estimation when the present estimation mechanism is applied to the network of vision cameras. Section V clarifies the tracking performance of the estimates when the target objects are moving. Verifications through simulation are shown in Section VI. Finally, Section VII draws conclusions.

We finally give some notations used in this paper, where the readers are recommended to refer to [3] for details on the terminologies. Throughout this paper, we use the notation $e^{\hat{\xi}_{ab}\theta_{ab}} \in \mathcal{R}^{3\times3}$ to represent the rotation matrix of a frame Σ_b relative to a frame Σ_a, which is orthogonal with unit determinant and hence an element of the Lie group $SO(3) := \{R \in \mathcal{R}^{3\times3}|\ R^T R = I_3 \text{ and } \det(R) = +1\}$. The vector $\xi_{ab} \in \mathcal{R}^3$ specifies the rotation axis and $\theta_{ab} \in \mathcal{R}$ is the rotation angle. For simplicity we use $\xi\theta_{ab}$ to denote $\xi_{ab}\theta_{ab}$. The configuration space of the rigid body motion is the product space $SE(3) := \mathcal{R}^3 \times SO(3)$. We use the 4×4 matrix $g_{ab} = \begin{bmatrix} e^{\hat{\xi}\theta_{ab}} & p_{ab} \\ 0 & 1 \end{bmatrix}$ as the homogeneous representation of $g_{ab} = (p_{ab}, e^{\hat{\xi}\theta_{ab}}) \in SE(3)$ describing the configuration of Σ_b relative to Σ_a. The notation '$\wedge$' is the operator such that $\hat{a}b = a \times b$ for the vector cross-product $\times$, i.e. $\hat{a}$ is a 3×3 skew-symmetric matrix. The vector space of all 3×3 skew-symmetric matrices is denoted by $so(3)$. The notation '$\vee$' denotes the inverse operator to '$\wedge$'. Similarly to the definition of $so(3)$, we define $se(3) := \{(v, \hat{\omega}) : v \in \mathcal{R}^3, \hat{\omega} \in so(3)\}$. In homogeneous representation, we write an element $V := (v, \omega)$ as $\hat{V} = \begin{bmatrix} \hat{\omega} & v \\ 0 & 0 \end{bmatrix}$.

II. PREPARATION FOR VISUAL SENSOR NETWORKS

Let us consider the situation where n vision cameras $\mathcal{V} := \{1, \cdots, n\}$ with communication and computation capability see a group of target objects $\{o_i\}_{i\in\mathcal{V}}$ (Fig. 1), where each vision camera $i \in \mathcal{V}$ captures object o_i on its image plane. Throughout this paper, we use the pinhole-type vision cameras with perspective projection [3] as in Fig. 2. Note however that all of the subsequent discussions are applicable to panoramic cameras through the modifications in [29].

In this paper, we address estimation of average motion of the objects $\{o_i\}_{i\in\mathcal{V}}$. The problem includes a scenario such that all the cameras see a common single target object but the pose consistent with vision data differs from camera to camera due to incomplete localization and parametric uncertainties. Under such a situation, averaging the contaminated poses is a way to improve estimation accuracy [20].
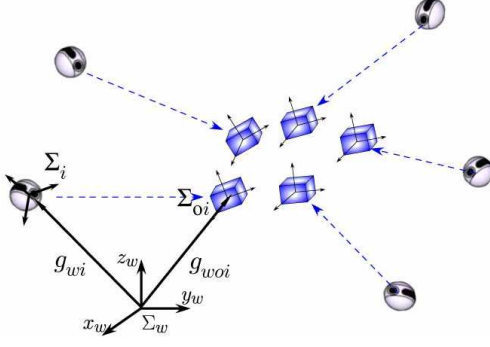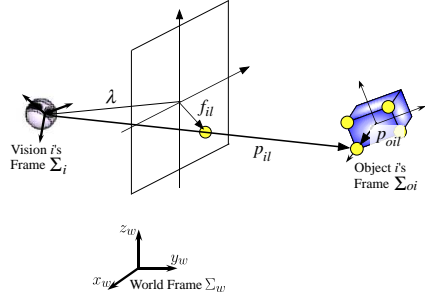
Fig. 1. Visual Sensor Networks

Fig. 2. Vision Camera Model

A. *Rigid Body Motion*

Let the coordinate frames Σ_w, Σ_i and Σ_{o_i} represent the world frame, the i-th vision camera frame, and the frame of object o_i, respectively. The pose of vision camera Σ_i and object Σ_{o_i} relative to the world frame Σ_w are denoted by $g_{wi} = (p_{wi}, e^{\hat{\xi}\theta_{wi}}) \in SE(3)$ and $g_{wo_i} = (p_{wo_i}, e^{\hat{\xi}\theta_{wo_i}}) \in SE(3)$. Then, the pose of Σ_{o_i} relative to Σ_i, denoted by $g_{io_i} = (p_{io_i}, e^{\hat{\xi}\theta_{io_i}}) \in SE(3)$, can be represented as $g_{io_i} = g_{wi}^{-1} g_{wo_i}$.

We next define the body velocity of object o_i relative to the world frame Σ_w as $V_{wo_i}^b = (v_{wo_i}, \omega_{wo_i}) \in \mathcal{R}^6$, where v_{wo_i} and ω_{wo_i} respectively represent the linear and angular velocities of the origin of Σ_{o_i} relative to Σ_w [3]. Similarly, vision camera i's body velocity relative to Σ_w will be denoted as $V_{wi}^b = (v_{wi}, \omega_{wi}) \in \mathcal{R}^6$.

By using the body velocities V_{wi}^b and $V_{wo_i}^b$, the motion of the relative pose g_{io_i} is written as

$$\dot{g}_{io_i} = -\hat{V}_{wi}^b g_{io_i} + g_{io_i} \hat{V}_{wo_i}^b \tag{1}$$

[3]. Equation (1) is called relative rigid body motion whose block diagram is depicted in Fig. 3.

B. *Visual Measurement*

In this subsection, we define visual measurements of each vision camera which is available for estimation. We assume that each target object has m feature points and each vision camera can extract them from the vision data by using some techniques like [30]. The position vectors of the target object i's l-th feature point relative to Σ_{o_i} and Σ_i are denoted by $p_{o_i l} \in \mathcal{R}^3$ and $p_{il} \in \mathcal{R}^3$ respectively. Using a transformation of the coordinates, we have $p_{il} = g_{io_i} p_{o_i l}$, where $p_{o_i l}$ and p_{il} should be regarded with a slight abuse of notation as $[p_{o_i l}^T \ 1]^T$ and $[p_{il}^T \ 1]^T$.
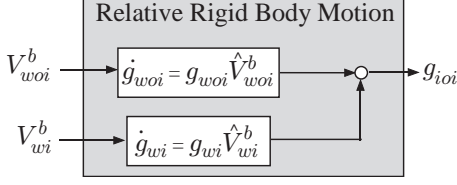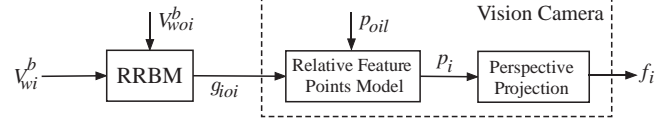
Fig. 3. Block Diagram of Relative Rigid Body Motion

Fig. 4. Block Diagram of the RRBM with Vision Camera (RRBM is an acronym for Relative Rigid Body Motion)

Let the m feature points of object o_i on the image plane coordinate be the measurement f_i of camera i, which is given by the perspective projection [3] with a focal length λ_i as

$$f_i := [f_{i1}^T \ \cdots \ f_{im}^T]^T \in \mathcal{R}^{2m}, \quad f_{il} = \frac{\lambda_i}{z_{il}}[x_{il} \ y_{il}], \ p_{il} = [x_{il} \ y_{il} \ z_{il}]^T, \tag{2}$$

Under the assumption that each camera i knows the location of feature points $p_{o_il} \in \mathcal{R}^3$, the visual measurement f_i depends only on the relative pose g_{io_i} from (2) and $p_{il} = g_{io_i}p_{o_il}$. Fig. 4 shows the block diagram of the relative rigid body motion with the camera model.

C. Communication

The vision cameras have communication capability with the neighboring cameras and form a network. The communication is modeled by a digraph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ as in the left figure of Fig. 5. Namely, vision camera i can get some information from j if $(j, i) \in \mathcal{E}$. In addition, we define the neighbor set $\mathcal{N}_i$ of vision camera $i \in \mathcal{V}$ as

$$\mathcal{N}_i := \{j \in \mathcal{V}|\ (j, i) \in \mathcal{E}\}. \tag{3}$$

Let us now employ the following assumption on the graph G.

Assumption 1: The communication graph G is fixed, balanced and strongly connected.

The balanced and strongly connected graph is a graph such that there exists at least one directed path between any pair of nodes and the in-degree and out-degree are equal for all nodes [11].

We also denote by G_u the undirected graph produced by replacing all the directed edges of G by the undirected ones. Let $\mathcal{T}(i_0)$ be the set of all spanning trees over G_u with a root $i_0 \in \mathcal{V}$ and we consider an element $G_T = (\mathcal{V}, \mathcal{E}_T) \in \mathcal{T}(i_0)$. Let the path from i_0 to a node $i \in \mathcal{V}$ along with the tree G_T be denoted by $P_{G_T}(i) = (v_0, \cdots, v_{d_{G_T}(i)}), \ v_0 = i_0, \ v_{d_{G_T}(i)} = i, \ (v_l, v_{l+1}) \in$
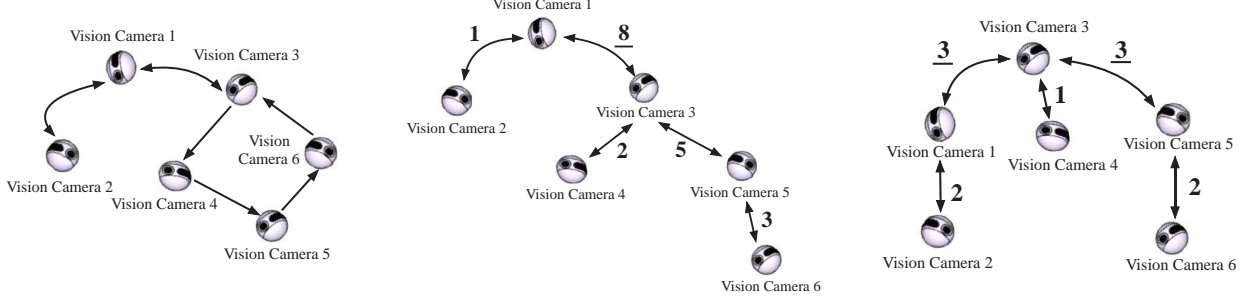
Fig. 5. Left: Communication Graph, Middle: Tree with Root 1 Minimizing $\tilde{D}$ ($\tilde{D} = 8$), Right: Tree Minimizing D ($W = 3$)

$\mathcal{E}_T$ $\forall l \in \{0, \cdots, d_{G_T}(i) - 1\}$, where $d_{G_T}(i)$ denotes the length of the path $P_{G_T}(i)$. We also define

$$\delta_{G_T}(E; i) = \begin{cases} 1, & \text{if the path } P_{G_T}(i) \text{ includes edge } E \\ 0, & \text{otherwise} \end{cases}$$

for any $E \in \mathcal{E}_T$. By using the above notations, we define

$$W := \min_{i_0 \in \mathcal{V}} D(i_0), \quad D(i_0) := \min_{G_T \in \mathcal{T}(i_0)} \tilde{D}(G_T), \quad \tilde{D}(G_T) := \max_{E \in \mathcal{E}_T} \sum_{i \in \mathcal{V}} \delta_{G_T}(E; i) d_{G_T}(i). \quad (4)$$

For example, let us consider the communication graph in Fig. 5(Left). Suppose that we choose $i_0 = 1$ and build a tree depicted in the middle figure of Fig. 5, where the number at around each edge is the value of $\sum_{i \in \mathcal{V}} \delta_{G_T}(E; i) d_{G_T}(i)$. Namely, $\tilde{D}$ is equal to 8 for the tree and it is actually minimal for all spanning trees in $\mathcal{T}(1)$. However, choosing another node as a root can reduce the value of $\tilde{D}$. Indeed, as illustrated in the right figure of Fig. 5, a tree with $i_0 = 3$ achieves $\tilde{D} = 3$, which is the minimal $D(i_0)$ among all the choices of the root i_0.

D. Average on $SO(3)$ and $SE(3)$

In this paper, the tuple of the relative rigid body motion (1), the visual measurement (2) and the communication structure (3) is called a visual sensor network. The objective of this paper is to present a cooperative estimation mechanism for the visual sensor networks meeting the following requirements simultaneously: *Averaging for static objects*, which means each camera i estimates a pose close to an average of $\{g_{io_j}\}_{j \in \mathcal{V}}$, $g_{io_j} := g_{wi}^{-1} g_{wo_j}$, *Tracking to moving objects*, which means the estimates track the moving average pose within a bounded tracking error.

Let us now introduce the following mean g^* on $SE(3)$ as an average of target poses $\{g_{wo_j}\}_{j \in \mathcal{V}}$.

$$g^* = (p^*, e^{\hat{\xi}\theta^*}) := \arg \min_{g \in SE(3)} \sum_{j \in \mathcal{V}} \psi(g^{-1} g_{wo_j}), \quad (5)$$

where the function ψ is defined for any $g = (p, e^{\hat{\xi}\theta}) \in SE(3)$ as

$$\psi(g) := \frac{1}{2}\|I_4 - g\|_F^2 = \frac{1}{2}\|p\|^2 + \phi(e^{\hat{\xi}\theta}), \quad \phi(e^{\hat{\xi}\theta}) := \frac{1}{2}\|I_3 - e^{\hat{\xi}\theta}\|_F^2 = \text{tr}(I_3 - e^{\hat{\xi}\theta}) \quad (6)$$

and $\|M\|_F$ is the matrix Frobenius norm of matrix M. Hereafter, we also use the notation

$$g_i^* = (p_i^*, e^{\hat{\xi}\theta_i^*}) := \arg\min_{g_i \in SE(3)} \sum_{j \in \mathcal{V}} \psi(g_i^{-1} g_{io_j}) = g_{wi}^{-1} g^*.$$

The position average p^* is equal to the arithmetic mean $p^* = \frac{1}{n}\sum_{j \in \mathcal{V}} p_{wo_j}$ of target positions $\{p_{wo_j}\}_{j \in \mathcal{V}}$ and the orientation average $e^{\hat{\xi}\theta^*}$ is a so-called Euclidean mean [25] of target orientations $\{e^{\hat{\xi}\theta_{wo_j}}\}_{j \in \mathcal{V}}$ defined by

$$e^{\hat{\xi}\theta^*} := \arg\min_{e^{\hat{\xi}\theta} \in SO(3)} \sum_{j \in \mathcal{V}} \phi(e^{-\hat{\xi}\theta} e^{\hat{\xi}\theta_{wo_j}}). \quad (7)$$

It is known [25] that the Euclidean mean $e^{\hat{\xi}\theta^*}$ is given by

$$e^{\hat{\xi}\theta^*}(t) = \text{Proj}(S(t)), \quad S(t) := \frac{1}{n}\sum_{j \in \mathcal{V}} e^{\hat{\xi}\theta_{wo_j}}(t). \quad (8)$$

Here, $\text{Proj}(M)$ is the orthogonal projection of $M \in \mathcal{R}^{3 \times 3}$ onto $SO(3)$, which is given by $U_M V_M^T$ for the matrix M with singular value decomposition $M = U_M \Sigma V_M^T$ [25].

Remark 1: Just computing the Euclidean mean is not so difficult even in a distributed fashion if we have prior knowledge that the target object is static. Indeed, the matrix S is computed by using the consensus protocol under appropriate assumptions on the graph [10] and the operation Proj can be locally executed. However, such a scheme works only for static objects and never embodies tracking nature for moving target objects. The objective here is to present an estimation mechanism without using any prior knowledge and any decision-making process on whether the targets are static or moving.

III. Networked Visual Motion Observer

A. Visual Motion Observer

In this subsection, we consider the problem that vision camera i estimates the target object motion g_{io_i} from the visual measurements f_i without considering communication. For the purpose, we introduce the *visual motion observer* presented in [8].

We first prepare a model of the rigid body motion (1) similarly to the Luenberger observer as

$$\dot{\bar{g}}_{io_i} = -\hat{V}_{wi}^b \bar{g}_{io_i} + \bar{g}_{io_i} \hat{u}_{ei}, \quad (9)$$
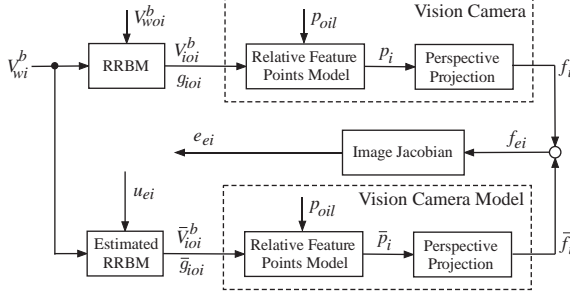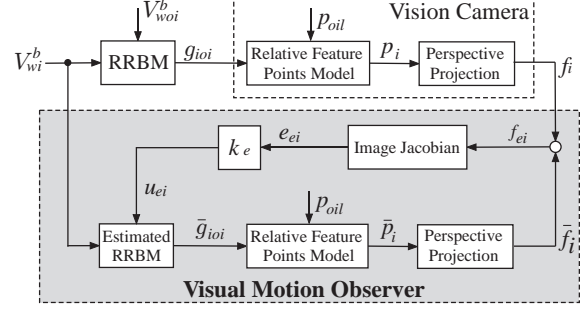
Fig. 6. Estimation Error System

Fig. 7. Visual Motion Observer

where $\bar{g}_{io_i} = (\bar{p}_{io_i}, e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})$ is the estimate of the actual relative pose g_{io_i}. The input $u_{ei} = (v_{uei}, \omega_{uei})$ is to be determined to drive the estimated value $\bar{g}_{io_i}$ to the actual g_{io_i}.

In order to establish the estimation error system, we define the estimation error between the estimated value $\bar{g}_{io_i}$ and the actual relative rigid body motion g_{io_i} as $g_{ei} = (p_{ei}, e^{\hat{\xi}\theta_{ei}}) := \bar{g}_{io_i}^{-1} g_{io_i}$. Using the notations $e_R(e^{\hat{\xi}\theta}) := \mathrm{sk}(e^{\hat{\xi}\theta})^{\vee}$ and $\mathrm{sk}(e^{\hat{\xi}\theta}) := \frac{1}{2}(e^{\hat{\xi}\theta} - e^{-\hat{\xi}\theta})$, the vector representation of the estimation error g_{ei} is given by

$$e_{ei} := E_R(g_{ei}),\ E_R(g_{ei}) := \left[\begin{array}{cc} p_{ei}^T & e_R^T(e^{\hat{\xi}\theta_{ei}}) \end{array}\right]^T. \tag{10}$$

Once the estimate $\bar{g}_{io_i}$ is determined, the estimated measurement $\bar{f}_i$ is also computed by (2). Let us now define the visual measurement error as $f_{ei} := f_i(g_{io_i}) - \bar{f}_i(\bar{g}_{io_i})$. Then, the measurement error vector f_{ei} can be approximately given by $f_{ei} = J_i(\bar{g}_{io_i})e_{ei}$ [8], where $J_i(\bar{g}_{io_i}) : SE(3) \to \mathcal{R}^{2m \times 6}$ is the well-known image Jacobian. Now, if $m \geq 4$, the image Jacobian has the full column rank and the estimation error vector e_{ei} is reconstructed as

$$e_{ei} = J_i^{\dagger}(\bar{g}_{io_i})f_{ei}, \tag{11}$$

where $\dagger$ denotes the pseudo-inverse.

Differentiating $g_{ei} = \bar{g}_{io_i}^{-1} g_{io_i}$ with respect to time and using (1) and (9), we obtain the estimation error system

$$\dot{g}_{ei} = -\hat{u}_{ei}g_{ei} + g_{ei}\hat{V}_{wo_i}^b. \tag{12}$$

Fig. 6 shows the block diagram of the system (12). The paper [8] proves that if $V_{wo_i}^b = 0$, then the estimation error system (12) is passive from the input u_{ei} to the output $-e_{ei}$.

Based on passivity-based control theory, we close the loop by using the input

$$u_{ei} = -k_e(-e_{ei}) = k_e e_{ei}, \ k_e > 0. \tag{13}$$

Then, the resulting total estimation mechanism formulated as

$$\text{Visual Motion Observer:} \begin{cases} \dot{\bar{g}}_{io_i} = -\hat{V}^b_{wi}\bar{g}_{io_i} + \bar{g}_{io_i}\hat{u}_{ei} & \cdots(9) \\ e_{ei} = J_i^{\dagger}(\bar{g}_{io_i})f_{ei} & \cdots(11) \\ u_{ei} = k_e e_{ei} & \cdots(13) \end{cases} \tag{14}$$

is called *visual motion observer* [9], whose block diagram is illustrated in Fig. 7. In terms of the mechanism, we immediately obtain the following facts from passivity.

Fact 1: [8] (i) If $V^b_{wo_i} = 0$, then the equilibrium point $e_{ei} = 0$ for the closed-loop system (12) with (13) is asymptotically stable. (ii) Given a positive scalar ν_i, if k_e satisfies $k_e - \frac{1}{2\nu_i^2} - \frac{1}{2} > 0$, then the system (12) and (13) with input $V^b_{wo_i}$ and output e_{ei} has L_2-gain smaller than ν_i. Item (i) means the visual motion observer leads the estimate $\bar{g}_{io_i}$ to the actual g_{io_i} for a static object. Item (ii) implies that the observer also works for a moving target object, and the parameter ν_i is an index on estimation accuracy when the observer is applied to a moving target.

B. Networked Visual Motion Observer

The objective of this paper is to achieve averaging, while preserving the tracking nature of the visual motion observer. For this purpose, this subsection presents a cooperative estimation mechanism under the assumption of (i) each vision camera knows relative pose $g_{ij} = g_{wi}^{-1}g_{wj}$ with respect to neighbors $j \in \mathcal{N}_i$ and (ii) all the vision cameras are static, i.e. $V^b_{wi} = 0 \ \forall i \in \mathcal{V}$.

Under $V^b_{wi} = 0$, the relative rigid body motion (1) is simply given by $\dot{g}_{io_i} = g_{io_i}\hat{V}^b_{wo_i}$. Accordingly, the update procedure in (14) is reformulated as

$$\dot{\bar{g}}_{io_i} = \bar{g}_{io_i}\hat{u}_{ei}, \ u_{ei} = k_e e_{ei}. \tag{15}$$

Then, the following proposition holds in terms of the procedure (15).

Proposition 1: [31] The update procedure (15) is a gradient decent algorithm on $SE(3)$ for the potential function $\psi(\bar{g}_{io_i}g_{io_i})$, where the function ψ is defined in (6).

Let us now view $\psi(\bar{g}_{io_i}^{-1}g_{io_i}) = \psi(\bar{g}_{wo_i}^{-1}g_{wo_i})$ as the local objective function to be minimized by vision camera i. Then, we see that the group objective (5) is given by the sum of the local objective functions for all $i \in \mathcal{V}$. Note that each vision camera does not know the local objective

Fig. 8. Networked Visual Motion Observer

of the other vision cameras. Under such a situation computing a solution minimizing the global objective function by using local negotiations is called *multi-agent optimization problem* and [32] presents an update rule of the local estimates of the solution to produce approximate solutions to the global objective combining the gradient decent algorithm of the local objective function and the consensus protocol [10]. The present cooperative estimation mechanism is inspired by the algorithm but the consensus protocol cannot be executed on $SE(3)$. We thus instead use a pose synchronization law presented in [15], which is also based on passivity of rigid body motion.

We next present an update rule of the estimates $\bar{g}_{io_i}$ so as to estimate the average g_i^*. Each vision camera i first gains the estimates $\bar{g}_{jo_j}$ from $j \in \mathcal{N}_i$ as messages. Now, by multiplying known information g_{ij} from left, each vision camera i gets $\bar{g}_{io_j} := g_{ij}\bar{g}_{jo_j}$ for all $j \in \mathcal{N}_i$. Using the information, the estimate $\bar{g}_{io_i}$ is updated according to (9) with

$$u_{ei} = k_e e_{ei} + k_s \sum_{j \in \mathcal{N}_i} E_R(\bar{g}_{io_i}^{-1}\bar{g}_{io_j}), \ k_e > 0, \ k_s > 0. \tag{16}$$

Since e_{ei} is reconstructed from the visual measurement f_i by (11) and $\bar{g}_{io_j}$ is obtained through communication as stated above, the update procedure (16) is implementable.

The present input (16) consists of the visual feedback term $k_e e_{ei}$ and the mutual feedback term $k_s \sum_{j \in \mathcal{N}_i} E_R(\bar{g}_{io_i}^{-1}\bar{g}_{io_j})$, where the former is inspired by the visual motion observer [8] and the latter is by the pose synchronization law [15]. Indeed, without the second term, the update rule (16) is the same as that of the visual motion observer (15). In addition, without the

visual feedback, the update procedure (16), namely $u_{ei} = k_s \sum_{j \in \mathcal{N}_i} E_R(\bar{g}_{io_i}^{-1} \bar{g}_{io_j})$, is essentially equivalent to the passivity-based pose synchronization law [15] of a group of rigid bodies with states $\bar{g}_{wo_i} := g_{wi} \bar{g}_{io_i}$. Thus, under appropriate assumptions, each state $\bar{g}_{wo_i}$ would converge to a state satisfying $\bar{g}_{wo_i} = \bar{g}_{wo_j}$ $\forall i, j \in \mathcal{V}$ as time goes to infinity without the visual feedback term.

In other words, the visual motion observers are networked by the mutual feedback term in the total estimation mechanism formulated as

$$\text{Networked VMO:} \begin{cases} \dot{\bar{g}}_{io_i} = \bar{g}_{io_i} \hat{u}_{ei} & \cdots (9) \\ e_{ei} = J_i^{\dagger}(\bar{g}_{io_i}) f_{ei} & \cdots (11) \quad \forall i \in \mathcal{V}, \\ u_{ei} = k_e e_{ei} + k_s \sum_{j \in \mathcal{N}_i} E_R(\bar{g}_{io_i}^{-1} \bar{g}_{io_j}) & \cdots (16) \end{cases} \quad (17)$$

where VMO is an acronym for Visual Motion Observer. This is why the estimation mechanism is called *networked visual motion observer*. The block diagram of the total system of vision camera i is illustrated in Fig. 8.

IV. AVERAGING PERFORMANCE ANALYSIS

In this section, we derive ultimate estimation accuracy of the average g_i^* achieved by the networked visual motion observer (17) under the following assumption.

Assumption 2:

(i) The target objects are static, i.e. $V_{wo_i}^b = 0$ $\forall i \in \mathcal{V}$.

(ii) There exists a pair $(i, j) \in \mathcal{V} \times \mathcal{V}$ such that $p_{wo_i} \neq p_{wo_j}$ and $e^{\hat{\xi}\theta_{wo_i}} \neq e^{\hat{\xi}\theta_{wo_j}}$.

(iii) $e^{-\hat{\xi}\theta_i^*} e^{\hat{\xi}\theta_{io_i}} > 0$ for all $i \in \mathcal{V}$. [1]

The moving target objects will be investigated in Section V. The item (ii) is assumed in order to avoid a meaningless problem such that $g_{wo_i} = g_{wo_j}$ $\forall i, j \in \mathcal{V}$. Indeed, under the situation, it is straightforward to prove convergence of the estimates to the common pose. In terms of the item (iii), we see that if $e^{-\hat{\xi}\theta_{wo_i}} e^{\hat{\xi}\theta_{wo_j}} > 0$ for all $i, j \in \mathcal{V}$, then the following inequality holds.

$$\phi\big(e^{-\hat{\xi}\theta_i^*} e^{\hat{\xi}\theta_{io_i}}\big) \leq \phi_m := \max_{i,j \in \mathcal{V}} \phi\big(e^{-\hat{\xi}\theta_{wo_i}} e^{\hat{\xi}\theta_{wo_j}}\big) \ \forall i \in \mathcal{V} \quad (18)$$

Inequality (18) implies that if $e^{-\hat{\xi}\theta_{wo_i}} e^{\hat{\xi}\theta_{wo_j}} > 0$ $\forall i, j \in \mathcal{V}$ (ϕ_m is smaller than 2), then (iii) is satisfied. Thus, (iii) can be checked if set-valued prior information on the target orientations, i.e. an upper bound of ϕ_m is available.

---

[1] Throughout this paper, we refer to a real matrix M, which is not necessarily symmetric, as a positive definite (positive semi-definite) matrix if and only if $x^T M x > 0$ ($x^T M x \geq 0$) for all nonzero vector x.

A. Definition of Averaging Performance

In this subsection, we introduce a notion of approximate averaging. For this purpose, we define the following sets for any positive parameter ε.

$$\Omega_p(\varepsilon) := \left\{ (\bar{p}_{io_i})_{i\in\mathcal{V}} \,\middle|\, \frac{1}{2}\sum_{i\in\mathcal{V}} \|\bar{p}_{io_i} - p_i^*\|^2 \leq \varepsilon\rho_p \right\}, \quad \rho_p := \frac{1}{2}\sum_{i\in\mathcal{V}} \|p_{io_i} - p_i^*\|^2 \tag{19}$$

$$\Omega_R(\varepsilon) := \left\{ (e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \,\middle|\, \sum_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) \leq \varepsilon\rho_R \right\}, \quad \rho_R := \sum_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\xi}\theta_{io_i}}) \tag{20}$$

Let us now define ε-level averaging performance to be met by the estimates $\bar{g}_{io_i} = (\bar{p}_{io_i}, e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})$.

Definition 1: Given target poses $(g_{io_i})_{i\in\mathcal{V}}$, position estimates $(\bar{p}_{io_i})_{i\in\mathcal{V}}$ are said to achieve ε-level averaging performance for a scalar $\varepsilon > 0$ if there exists a finite T such that $(\bar{p}_{io_i}(t))_{i\in\mathcal{V}} \in \Omega_p(\varepsilon) \ \forall t \geq T$ and the orientation estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ are said to achieve ε-level averaging performance if there exists a finite T such that $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(t))_{i\in\mathcal{V}} \in \Omega_R(\varepsilon) \ \forall t \geq T$.

In the absence of communication, each vision camera i acquires no information on the target objects o_j, $j \neq i$. Under the situation, what each vision camera can do is to produce as an accurate estimate of the relative pose g_{io_i} as possible. Namely, the parameters ρ_p and ρ_R specify the best performance of average estimation in the absence of communication. More specifically, since the visual motion observer (14) correctly estimates the static target object pose g_{io_i} (Fact 1), the parameters ρ_p and ρ_R indicate the average estimation accuracy in the absence of the mutual feedback term of u_{ei} in (16). Namely, the parameter ε is an indicator of improvement of average estimation accuracy by inserting the mutual feedback term $k_s \sum_{j\in\mathcal{N}_i} E_R(\bar{g}_{io_i}^{-1}\bar{g}_{io_j})$.

B. Auxiliary Results

In this subsection, we give some results necessary for proving the main result of this section.

Lemma 1: Suppose that the estimates $(\bar{g}_{io_i})_{i\in\mathcal{V}}$ are updated by the networked visual motion observer (17). Then, under Assumptions 1 and 2 and $e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0 \ \forall t \geq 0$, for all $c > 0$, there exists a finite $\tau(c)$ such that $\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) \leq \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}}) + c \ \forall t \geq \tau(c), \ i \in \mathcal{V}$, where $h := \arg\max_{j\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_j}})$.

Proof: See Appendix A. ∎

Lemma 1 implies that the individual estimate $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ gets closer to the average $e^{\hat{\xi}\theta_i^*}$ at least than the object with the farthest orientation from the average. In addition, the proof of this lemma also means that the set

$$\mathcal{S} = \{(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}|\ e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0\ \forall i \in \mathcal{V}\}$$

is positively invariant for the total system (17) under Assumption 2. Namely, if $e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0$ is satisfied at the initial time, then $e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0$ holds for all subsequent time.

We next have the following lemma.

Lemma 2: Suppose that the estimates $(\bar{g}_{io_i})_{i\in\mathcal{V}}$ are updated by the networked visual motion observer (17). Then, under Assumptions 1 and 2, if the initial estimates satisfy $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(0))_{i\in\mathcal{V}} \in \mathcal{S}$, both of the estimates $(\bar{p}_{io_i})_{i\in\mathcal{V}}$ and $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ achieve 1-level averaging performance.

Proof: See Appendix B. ∎

This lemma is proved by using the energy functions

$$U_p := \frac{1}{2}\sum_{i\in\mathcal{V}}\|p_i^* - \bar{p}_{io_i}\|^2 = \frac{1}{2}\sum_{i\in\mathcal{V}}\|p^* - \bar{p}_{wo_i}\|^2,\ U_R := \sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) = \sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})$$

which are defined by the sum of individual error between the average and the estimate. The functions $U_p \geq 0$ and $U_R \geq 0$ are equal to 0 if and only if $\bar{p}_{io_i} = p_i^*$ and $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}} = e^{\hat{\xi}\theta_i^*}\ \forall i \in \mathcal{V}$ respectively. The selection of the energy function is inspired by one of our previous works on pose synchronization [15] whose framework is originally presented in [13].

Lemma 2 means that the average estimation as a group in the presence of communication is at least more accurate than the case in the absence of communication. However, this lemma does not say how accurate estimates of the average the networked visual motion observer produces.

From Lemmas 1 and 2, the estimates $(\bar{p}_{io_i})_{i\in\mathcal{V}}$ and $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ settle into $\Omega_p(1)$ and $\mathcal{S}_1^R := \mathcal{S}\cap\Omega_R(1)$ in finite time, respectively. Let us now define the following subsets of $\Omega_p(1)$ and $\mathcal{S}_1^R$.

$$\mathcal{S}_2^p(k) := \left\{(\bar{p}_{io_i})_{i\in\mathcal{V}} \in \Omega_p(1)\Big|\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\frac{1}{2}\|\bar{p}_{wo_i} - \bar{p}_{wo_j}\|^2 \geq k\rho_p\right\},$$

$$\mathcal{S}_2^R(k) := \left\{(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \in \mathcal{S}_1^R\Big|\beta\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \geq k\rho_R\right\},$$

$$\mathcal{S}_3^p(k,\varepsilon) := \Omega_p(1) \setminus (\mathcal{S}_2^p(k)\cup\Omega_p(\varepsilon)),\ \mathcal{S}_3^R(k,\varepsilon) := \mathcal{S}_1^R \setminus (\mathcal{S}_2^R(k)\cup\Omega_R(\varepsilon))$$

for some $\varepsilon \in [0,1)$, where $\beta := 1 - \sqrt{2(\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}}) + c)}$ and $k = k_e/k_s$. Images of the subsets on the position space are depicted in Fig. 9. We see from the figure that

$$\Omega_p(1) \setminus (\mathcal{S}_2^p(k)\cup\mathcal{S}_3^p(k,\varepsilon)) \subseteq \Omega_p(\varepsilon),\ \mathcal{S}_1^R \setminus (\mathcal{S}_2^R(k)\cup\mathcal{S}_3^R(k,\varepsilon)) \subseteq \Omega_R(\varepsilon). \tag{21}$$
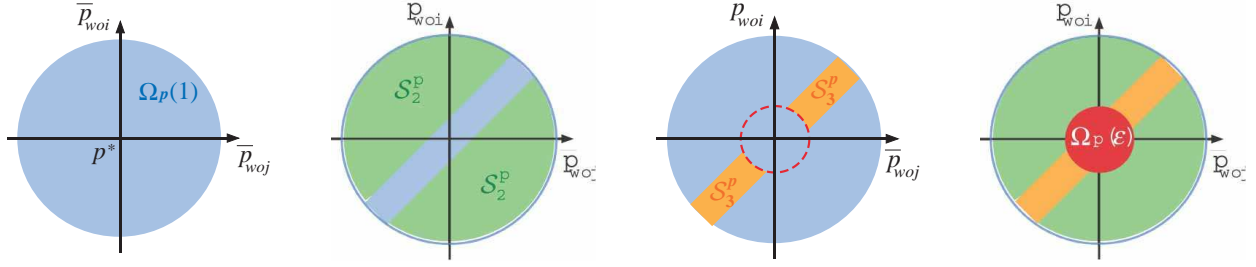
Fig. 9.   Images of Each Subsets (Position)

In terms of the subsets $\mathcal{S}_2^p(k)$ and $\mathcal{S}_2^R(k)$, we have the following lemma.

Lemma 3: Suppose that all the assumptions in Lemma 2 hold and $\beta > 0$. Then, the time derivative of U_p and U_R along with the trajectories of (17) are strictly negative as long as $(\bar{p}_{io_i})_{i\in\mathcal{V}} \in \mathcal{S}_2^p(k)$ and $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \in \mathcal{S}_2^R(k)$ respectively, at least after the time $\tau(c)$.

Proof: See Appendix C.   ∎

From (18), β can be estimated by set-valued prior information on the target orientations i.e. ϕ_m.

C. Averaging Performance

We are now ready to state the main result of this section on averaging accuracy attained by the networked visual motion observer (17).

Theorem 1: Suppose that all the assumptions in Lemma 2 hold. Then, for any $\epsilon \in (0,1)$, position estimates $(\bar{p}_{io_i})_{i\in\mathcal{V}}$ achieve ε_p-level averaging performance with

$$\varepsilon_p = \begin{cases} 1 - (1-\epsilon)\left(1 - \sqrt{kW}\right)^2 & \text{if } k \le 1/W \\ 1 & \text{otherwise} \end{cases}, \tag{22}$$

and orientation estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ achieve ε_R-level averaging performance with

$$\varepsilon_R = \begin{cases} 1 - (1-\epsilon)\left(\sqrt{\beta} - \sqrt{kW}\right)^2, & \text{if } k \le \beta/W,\ \beta > 0 \\ 1, & \text{otherwise} \end{cases}, \tag{23}$$

where W is defined in (4).

Proof: See Appendix D.   ∎

Suppose that ϵ is taken sufficiently close to 0. Then, we see that both of the parameters ε_p and ε_R become small as the term $\sqrt{kW}$ approaches to 0. Note that if we use a sufficiently small k ($k_s \gg k_e$) in (16), the term is approximated by 0. Here, we see an essential difference between

the position and orientation estimates. The definition of ε_p with $\epsilon \approx 1$ indicates that we can get arbitrarily accurate estimation of the average p_i^* by choosing a sufficiently small k. In contrast, we see from the definition of ε_R that an offset associated with $\sqrt{\beta}(< 1)$ occurs for the orientation estimates regardless of the parameter k. From the definition of $\beta := 1 - \sqrt{2(\phi(e^{-\hat{\xi}\theta^*} e^{\hat{\xi}\theta_{wo_h}}) + c)}$, if the target object's orientation $e^{\hat{\xi}\theta_{wo_h}}$ is sufficiently close to the average $e^{-\hat{\xi}\theta^*}$, i.e. if $e^{\hat{\xi}\theta_{wo_i}}$ and $e^{\hat{\xi}\theta_{wo_j}}$ are close among all $i, j \in \mathcal{V}$ enough to approximate all the orientations by matrices on a tangent vector space of $SO(3)$ at $e^{\hat{\xi}\theta_{wo_i}}$, then it becomes close to 0 and the average is accurately estimated by the networked visual motion observer (17). Otherwise, the accuracy might degrade, though it is more accurate at least than the case in the absence of communication.

V. TRACKING PERFORMANCE ANALYSIS

In this section, we analyze the tracking performance of the estimates $\{\bar{g}_{io_i}\}_{i \in \mathcal{V}}$ to the average g_i^* for moving targets when the networked visual motion observer is applied to the visual sensor networks under the following assumption.

Assumption 3:

(i) The target body velocities $V_{wo_i}^b(t)$, $i \in \mathcal{V}$ are continuous in t and bounded as

$$\|v_{wo_i}^b(t)\|_2^2 \leq \bar{w}_p^2, \ \|\omega_{wo_i}^b(t)\|^2 \leq \bar{w}_R^2 \ \forall i \in \mathcal{V}, \ t \geq 0. \tag{24}$$

(ii) For all $t \geq 0$, there exists $(i(t), j(t)) \in \mathcal{V} \times \mathcal{V}$ such that $p_{wo_{i(t)}} \neq p_{wo_{j(t)}}$ and $e^{\hat{\xi}\theta_{wo_{i(t)}}} \neq e^{\hat{\xi}\theta_{wo_{j(t)}}}$.

(iii) $e^{-\hat{\xi}\theta_{wo_j}}(t) e^{\hat{\xi}\theta_{wo_i}}(t) > 0$ for all $i, j \in \mathcal{V}$ and $t \geq 0$.

A. Description of Average Motion

In this subsection, we first formulate the motion of the average $g^* = (p^*, e^{\hat{\xi}\theta^*})$. The behavior of the position average p^* is clearly described by

$$\dot{p}^* = e^{\hat{\xi}\theta^*} v^{b,*}, \ v^{b,*} := e^{-\hat{\xi}\theta^*}\left(\frac{1}{n}\sum_{i \in \mathcal{V}} e^{\hat{\xi}\theta_{wo_i}} v_{wo_i}^b\right) \tag{25}$$

from the definition of $p^* = \frac{1}{n}\sum_{i \in \mathcal{V}} p_{wo_i}$. Meanwhile, the trajectory of the orientation average $e^{\hat{\xi}\theta^*}$ described by (8) satisfies the following lemma.

Lemma 4: Under Assumption 3, the average $e^{\hat{\xi}\theta^*}$ is continuously differentiable.

Proof: From the polar decomposition, we get $S(t) = e^{\hat{\xi}\theta^*}(t)P_S(t)$ [25], where $S(t) = \frac{1}{n}\sum_{i \in \mathcal{V}} e^{\hat{\xi}\theta_{wo_i}}$ and $P_S^2(t) = S^T(t)S(t)$. Under Assumption 3(iii), we have $e^{-\hat{\xi}\theta_{wo_j}}(t)S(t) > 0$ and

hence $P_S(t)$ is invertible for all $t \geq 0$. Thus, the average $e^{\hat{\xi}\theta^*}$ is given by $e^{\hat{\xi}\theta^*}(t) = S(t)P_S^{-1}(t)$. From (1), the matrices $S(t)$ and $P_S(t)$ are clearly differentiable from their definitions and hence $\dot{e}^{\hat{\xi}\theta^*}$ is well defined. Moreover, from Assumption 3(i), both of $\dot{S}(t)$ and $\dot{P}_S(t)$ are continuous and $\frac{d}{dt}(P_S^{-1}) = P_S^{-1}(t)\dot{P}_S P_S^{-1}$ is also continuous, which implies that $\dot{e}^{\hat{\xi}\theta^*}(t)$ is also continuous. Hence, the average $e^{\hat{\xi}\theta^*}$ is continuously differentiable. This completes the proof. ■

Moreover, since $e^{\hat{\xi}\theta^*}(t) \in SO(3)$ holds for all $t \geq 0$, the derivative $\dot{e}^{\hat{\xi}\theta^*}$ has to satisfy $\dot{e}^{\hat{\xi}\theta^*} \in T_{e^{\hat{\xi}\theta^*}}SO(3)$, where $T_{e^{\hat{\xi}\theta^*}}SO(3) := \{e^{\hat{\xi}\theta^*}X| \ X \in so(3)\}$ is the tangent space of the manifold $SO(3)$ at $e^{\hat{\xi}\theta^*}$. Namely, the trajectory of the Euclidean mean $e^{\hat{\xi}\theta^*}$ is described by the differential equation $\dot{e}^{\hat{\xi}\theta^*} = e^{\hat{\xi}\theta^*}\hat{\omega}^{b,*}$ with some body velocity $\hat{\omega}^{b,*} \in so(3)$.

We next clarify a relation between velocities $V^{b,*} := (v^{b,*}, \omega^{b,*})$ and $V^b_{wo_i} = (v^b_{wo_i}, \omega^b_{wo_i})$, $i \in \mathcal{V}$. We first define $w_p := (v^b_{wo_i})_{i \in \mathcal{V}}$ and $w_R := (\omega^b_{wo_i})_{i \in \mathcal{V}}$. Since it is easy from (25) to obtain $\|v^{b,*}\|^2 \leq \|w_p\|^2/n$, we mention only a relation between $\omega^{b,*}$ and w_R in the following.

Lemma 5: Suppose that the target orientations $(e^{\hat{\xi}\theta_{wo_i}})_{i \in \mathcal{V}}$ satisfy

$$\left\| e^{\hat{\xi}\theta^*}(t) - S(t) \right\|_F \leq \gamma \ \forall t \geq 0 \tag{26}$$

for some $\gamma > 0$. Then, the following inequality holds.

$$\|\omega^{b,*}(t)\|^2 < \frac{\mu^2(\gamma)}{n}\|w_R(t)\|^2, \ \mu(\gamma) := \frac{\sqrt{2}}{\sqrt{2} - \gamma} \tag{27}$$

Proof: See Appendix E ■

Though we omit the proof, $\left\| e^{\hat{\xi}\theta^*}(t) - S(t) \right\|_F$ is also upper bounded by ϕ_m and hence γ is estimated by prior information on the target orientations.

B. Tracking Performance

Let us consider the whole networked system Σ_{track} consisting of the relative rigid body motion (1) for all $i \in \mathcal{V}$ and the networked visual motion observer (17). Here, we regard the collections of body velocities of the target objects $(V^b_{wo_i})_{i \in \mathcal{V}}$, i.e. $w = (w_p, w_R)$, as the external disturbance to Σ_{track} and evaluate the error between the estimates $(\bar{g}_{io_i})_{i \in \mathcal{V}}$ and the average g^*_i in the presence of the disturbance w. Namely, we let the error $(\{g^*_i\}^{-1}\bar{g}_{io_i})_{i \in \mathcal{V}}$ be the output signal of Σ_{track}.

Unlike the static objects case, $\rho_p = \frac{1}{2}\sum_{i \in \mathcal{V}}\|p_{io_i} - p^*_i\|^2$ and $\rho_R = \sum_{i \in \mathcal{V}}\phi(e^{-\hat{\xi}\theta^*_i}e^{\hat{\xi}\theta_{io_i}})$ are also time-varying. We thus define the parameters

$$\rho'_p := \sup_t \rho_p(t), \ \rho'_R := \sup_t \rho_R(t)$$

assuming $\rho_p' < \infty$ and redefine the sets Ω_p' and Ω_R' by just replacing ρ_p and ρ_R in (19) and (20) by ρ_p' and ρ_R', respectively. The parameters ρ_p' and ρ_R' are the suprimum of the distance from the estimate to the average when g_{io_i} is correctly estimated and hence they are also indicators of the best average estimation performance in the absence of communication. Note however that the visual motion observer (14) cannot correctly estimate g_{io_i} as long as the object is moving with unknown velocity.

The problem to be considered here is redefined as follows.

Definition 2: The position estimates $(\bar{p}_{io_i})_{i\in\mathcal{V}}$ are said to achieve ε-level tracking performance for a positive scalar ε if there exists a finite T such that $(\bar{p}_{io_i}(t))_{i\in\mathcal{V}} \in \Omega_p'(\varepsilon)$ $\forall w \in \mathcal{W}$ and $t \geq T$, where $\mathcal{W}$ is the set of the disturbance signal $w(\cdot)$ satisfying Assumption 3. Similarly, the estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ are said to achieve ε-level tracking performance if there exists a finite T such that $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(t))_{i\in\mathcal{V}} \in \Omega_R'(\varepsilon)$ $\forall w \in \mathcal{W}$ and $t \geq T$.

In terms of the tracking performance defined above, we have the following theorem.

Theorem 2: Suppose that the estimates $\bar{g}_{io_i}$ are updated according to (17). Under Assumptions 1 and 3, if $k_e > \mu^2(\gamma)$ for γ satisfying (26) and $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(t))_{i\in\mathcal{V}} \in \mathcal{S}$ $\forall t \geq 0$, the estimates $(\bar{p}_{io_i})_{i\in\mathcal{V}}$ and $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ achieve ε_p' and ε_R'-level tracking performances respectively with

$$\varepsilon_p' := 1 + \frac{1}{k_e - 1} + \frac{\bar{w}_p^2}{\rho_p'(k_e - 1)}, \ \ \varepsilon_R' := 1 + \frac{\mu^2(\gamma)}{k_e - \mu^2(\gamma)} + \frac{\bar{w}_R^2}{\rho_R'(k_e - \mu^2(\gamma))}.$$

Proof: See Appendix F. ∎

This theorem implies that the networked visual motion observer works even for moving target objects. We also see that the ultimate error between the estimates and the average gets small as the visual feedback gain k_e becomes large, which is a natural conclusion from the form of (16).

In summary, we have the following conclusion on the gain selection. In order to achieve a good averaging performance, we should make the mutual feedback gain k_s large relative to the visual feedback gain k_e. In order to achieve a good tracking performance, the visual feedback gain k_e should be absolutely large. Namely, the best selection is to make both gains k_e and k_s large while the mutual feedback gain k_s is much larger than the visual feedback gain k_e. However, the size of k_s is in general restricted by the communication rate due to limitation in standard feedback control theory. Then, a trade-off occurs between averaging and tracking performances, i.e. if we set a large k_e, a good tracking performance is achieved at the cost of a poor averaging performance and vice versa.
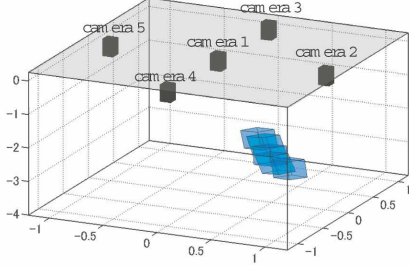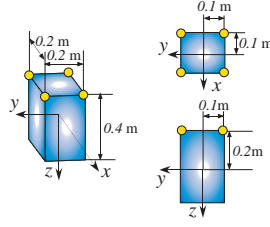
Fig. 10.   Overview
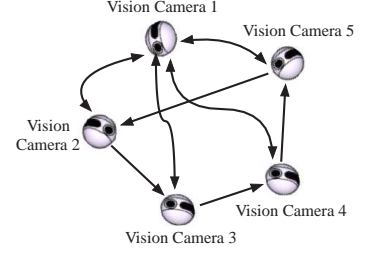

Fig. 11.   Feature Points

Fig. 12.   Communication Graph

VI. SIMULATION

We finally demonstrate the effectiveness of the networked visual motion observer and validity of the theoretical results through simulation. Throughout this section, we consider the situation where five pin-hole type vision cameras with focal length 0.01[m] see a group of target objects. We identify the frame of camera 1 with the world frame and let $p_{w2} = [1 \ 0 \ 0]^T$, $p_{w3} = [0 \ 1 \ 0]^T$, $p_{w4} = [0 \ -1 \ 0]^T$, $p_{w5} = [-1 \ 0 \ 0]^T$, and $e^{\hat{\xi}\theta_{wi}} = I_3$, $\forall i \in \{2, 3, 4, 5\}$. The overview of the setting is illustrated in Fig. 10, where blue boxes represent the initial configuration of target objects with $p_{wo_1} = [0.12 \ 0.55 \ -2.78]$, $p_{wo_2} = [0.22 \ 0.48 \ -2.85]$, $p_{wo_3} = [0.33 \ 0.33 \ -2.97]$, $p_{wo_4} = [0.42 \ 0.23 \ -3.08]$, $p_{wo_5} = [0.56 \ 0.12 \ -3.15]$ and $\xi\theta_{wo_1} = [-0.30 \ -0.30 \ -0.30]$, $\xi\theta_{wo_2} = [-0.30 \ -0.40 \ -0.40]$, $\xi\theta_{wo_3} = [-0.40 \ -0.30 \ -0.30]$, $\xi\theta_{wo_4} = [-0.30 \ -0.40 \ -0.30]$, $\xi\theta_{wo_5} = [-0.30 \ -0.30 \ -0.40]$. All the targets have four feature points whose positions relative to the object frame are illustrated in Fig. 11. We use the points projected onto the image plane as visual measurements f_i. The communication structure is depicted in Fig. 12 with $W = 1$.

In the first scenario, we consider static target objects and demonstrate validity of Theorem 1. Then, the average $g^* = (p^*, e^{\hat{\xi}\theta^*})$ is given by $p^* = [0.33 \ 0.36 \ -2.96]$, $\xi\theta^* = [-0.32 \ -0.34 \ -0.34]$. For the configuration of the target objects, the parameter β is given by about $\beta = 0.86$. Throughout this section, we let the initial estimates be $\bar{p}_{io_i}(0) = [0 \ 0 \ 2.5]^T$ and $e^{\hat{\xi}\bar{\theta}_{io_i}}(0) = I_3$.

We first employ the gains $k_e = 1$ and $k_s = 0.1$ ($k = 10$). Then, the parameters ε_p and ε_R in Theorem 1 are given by $\varepsilon_p = \varepsilon_R = 1$. Fig. 13 illustrates the time responses of orientation estimates of all vision cameras produced by the networked visual motion observer, where the red dash-dotted lines represent each element of the average $\xi^* \sin \theta^*$. We see from the figures that
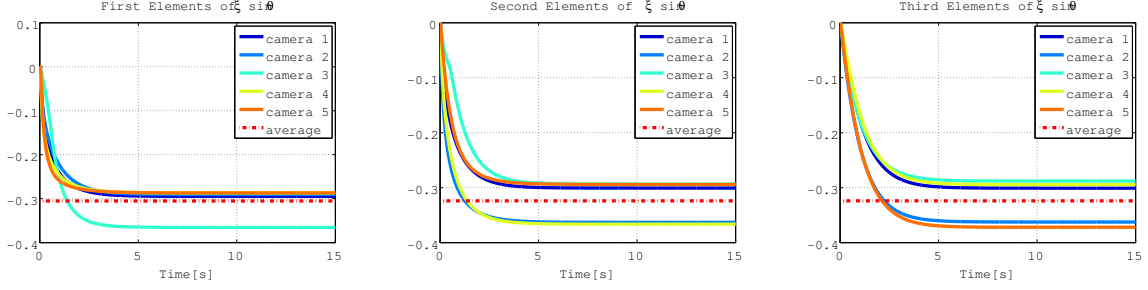
Fig. 13.   Time Responses of Each Element of $\bar{\xi} \sin \bar{\theta}$, $i = 1, \cdots, 5$ (Static: $k_s = 0.1$)
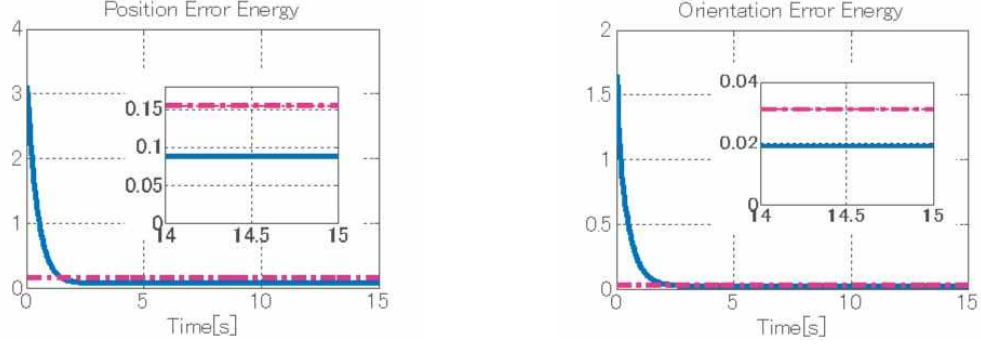

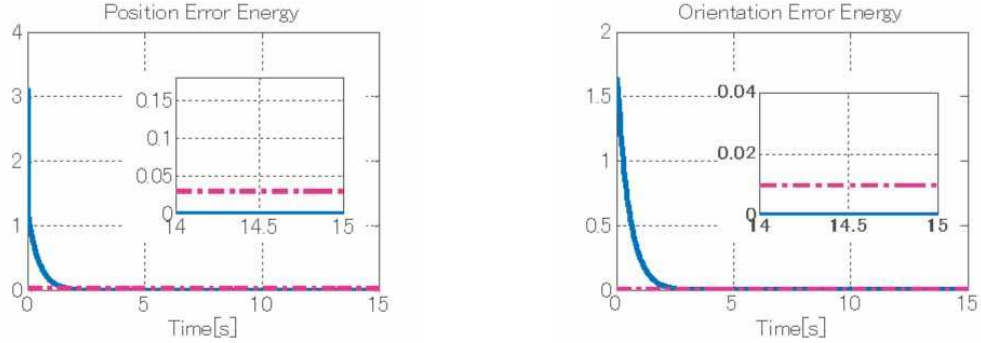Fig. 14.   Time Response of U_p (Left) and U_R (Right) (Static: $k_s = 0.1$)

Fig. 15.   Time Response of U_p (Left) and U_R (Right) (Static: $k_s = 100$)

there exist gaps between the average and the estimates for all elements. The error functions U_p and U_R are depicted by blue curves in Fig. 14 respectively, where red dash dotted lines represent $\varepsilon_p \rho_p$ and $\varepsilon_R \rho_R$. Namely, Theorem 1 implies that the blue curve eventually takes lower values than the value indicated by the dash dotted line and we see that it is really achieved as expected.

We next let $k_e = 1$ and $k_s = 100$ ($k = 0.01$). Then, we have $\varepsilon_p = 0.19$, $\varepsilon_R = 0.31$ for sufficiently small ϵ and c. Fig. 15 illustrates the time responses of U_p and U_R. We see from the figures that the estimates of all vision cameras become much closer to the average than the case
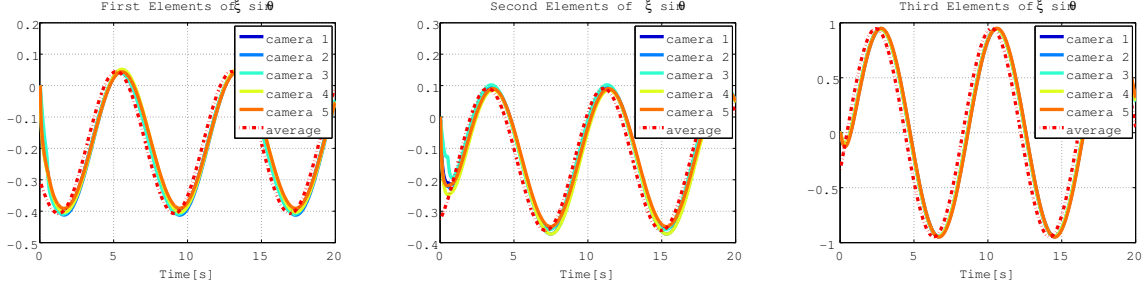

Fig. 16. Time Responses of Each Element of $\bar{\xi}\sin\bar{\theta}$, $i = 1, \cdots, 5$ (Moving: $k_e = 3$)
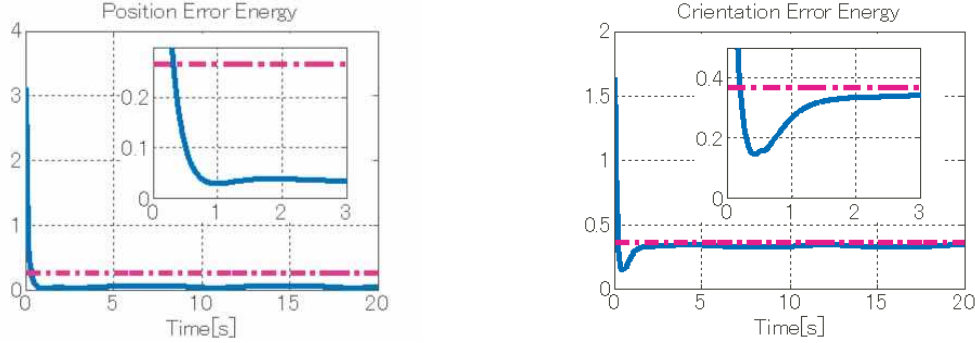

Fig. 17. Time Responses of U_p (Left) and U_R (Right) (Moving: $k_e = 3$)
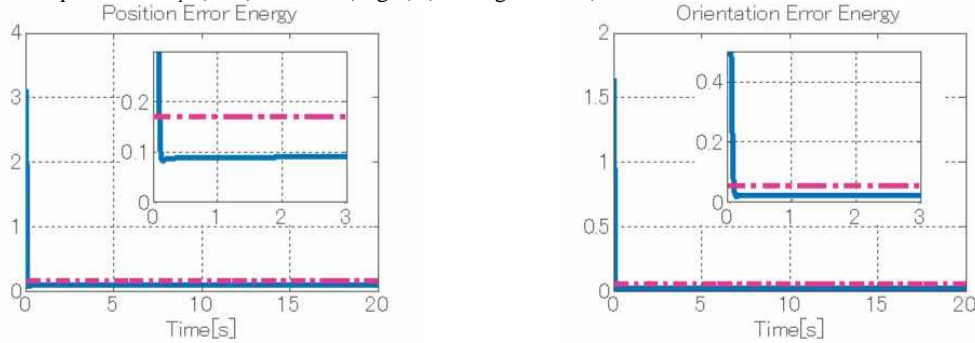

Fig. 18. Time Responses of U_p (Left) and U_R (Right) (Moving: $k_e = 30$)

of a small mutual feedback gain $k_s = 0.1$. Fig. 15 also indicates that the error functions U_p and U_R ultimately take lower values than the right-hand side of (22) and (23) respectively. Namely, it turns out as predicted that a small $k = k_e/k_s$ results in a good averaging performance.

In the second scenario, we consider moving target objects with constant body velocities $V_{wo_i}^b = \begin{bmatrix} 0.2 & 0 & 0 & 0 & 0 & 0.8 \end{bmatrix}^T$ $\forall i = 1, \cdots, 5$ and the same initial states as the above static case. For the targets, we apply the networked visual motion observer with $k_e = k_s = 3$, where we let the initial estimates be the same as the above static object case. Then the time responses of orientation estimates are depicted in Fig. 16, where red dash dotted curves describe the average

motion of the target orientations. We see from the figures that the estimates track the moving average within bounded errors and the networked observer also works for a dynamic problem.

The responses of U_p and U_R are illustrated in Fig. 17, where the dash-dotted lines show $\varepsilon'_p \rho'_p$ and $\varepsilon'_R \rho'_R$. As shown in Theorem 2, both of U_p and U_R ultimately take values smaller than $\varepsilon'_p \rho'_p$ and $\varepsilon'_R \rho'_R$ respectively. Their counterparts for $k_e = 30$, $k_s = 3$ are shown in Fig. 18, which also illustrate validity of Theorem 2. We also see that a large k_e achieves a better tracking performance than a smaller k_e, which supports validity of the analysis at the end of Section V.

Experimental verifications on a testbed are omitted in this paper but shown in [27], [28].

VII. CONCLUSIONS

This paper has presented a novel cooperative estimation mechanism for visual sensor networks. We have considered the situation where multiple smart vision cameras with computation and communication capability see a group of target objects. We first have presented an estimation mechanism called networked visual motion observer to meet two requirements, averaging and tracking. Then, we have derived an upper bound of the ultimate error between the actual average and the estimates produced by the present methodology. Moreover, we have derived an upper bound of the ultimate error from the estimates to the average when the target objects are moving. Finally, the effectiveness of the present mechanism has been demonstrated through simulation.

The authors would like to express sincere appreciation to Prof. Francesco Bullo and Prof. Kenji Hirata for their invaluable suggestions and advices.

APPENDIX A
PROOF OF LEMMA 1

In the proof, we use the following lemma.

Lemma 6: [15] For any matrices $R_1, R_2, R_3 \in SO(3)$, the inequality

$$\frac{1}{2}\mathrm{tr}(R_1^T R_2 - R_1^T R_3 R_2^T R_3) \geq \phi(R_1^T R_3) - \phi(R_1^T R_2) + \lambda_{min}(\mathrm{sym}(R_1^T R_3))\phi(R_3^T R_2)$$

holds, where $\mathrm{sym}(M) := \frac{1}{2}(M + M^T)$ and $\lambda_{min}(M)$ is the minimal eigenvalue of matrix M.

The time evolution of the orientation estimate $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ in (9) with $V_{wi}^b = 0$ and (16) is given by

$$\dot{e}^{\hat{\bar{\xi}}\bar{\theta}_{io_i}} = e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}\hat{\omega}_{uei}, \quad \omega_{uei} = k_e e_R(e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_{io_i}}) + k_s \sum_{j \in \mathcal{N}_i} e_R(e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{io_j}}), \tag{28}$$

which is independent of evolution of the position estimate $\bar{p}_{io_i}$. Multiplying $e^{\hat{\xi}\theta_{wi}}$ to (28) from left, we have the following equation describing evolution of the estimate $e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}$ relative to Σ_w.

$$\dot{e}^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} = e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}\hat{\omega}_{uei}, \ \ \omega_{uei} = k_e e_R(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\xi}\theta_{wo_i}}) + k_s \sum_{j\in\mathcal{N}_i} e_R(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \tag{29}$$

Let us now consider the energy function

$$U := \phi(e^{-\hat{\xi}\theta_{\iota}^*}e^{\hat{\bar{\xi}}\bar{\theta}_{\iota o_\iota}}) = \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}), \ \ \iota(t) := \arg\max_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(t)).$$

Then, the time derivative of U along with the trajectories of (29) is given by

$$\dot{U} = 2e_R^T(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}})\omega_{ue\iota} = -\mathrm{tr}\left(\mathrm{sk}(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_\iota}})\hat{\omega}_{ue\iota}\right), \tag{30}$$

where we use the relation $a^T b = -\frac{1}{2}\mathrm{tr}(\hat{a}\hat{b})$. Substituting (29) into (30) yields

$$\dot{U} = -\mathrm{tr}\Big(k_e\Big(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_\iota}} - e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}e^{-\hat{\xi}\theta_{wo_\iota}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}\Big)$$
$$+k_s\Big(\sum_{j\in\mathcal{N}_\iota} e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}} - e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_j}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}\Big)\Big). \tag{31}$$

From Lemma 6, (31) is rewritten as $\dot{U} \leq -(k_e F_1 + k_s F_2)$, where

$$F_1 := \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}) - \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_\iota}}) + \sigma\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}e^{\hat{\xi}\theta_{wo_\iota}}), \ \ \sigma := \lambda_{min}(\mathrm{sym}(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}})),$$
$$F_2 := \sum_{j\in\mathcal{N}_\iota}\left(\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}) - \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) + \sigma\phi(e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}})\right).$$

From the definition of the index ι, the inequality $\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}) \geq \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \ \forall j\in\mathcal{V}$ holds and hence we obtain $F_2 \geq \sigma\sum_{j\in\mathcal{N}_i}\phi(e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}})$. Thus, the inequality

$$\dot{U} \leq -\Big(k_e\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}) - k_e\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_\iota}}) + \sigma\Big(k_e\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}e^{\hat{\xi}\theta_{wo_\iota}}) + k_s\sum_{j\in\mathcal{N}_i}\phi(e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}})\Big)\Big).$$

is true. From the assumption of $e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} > 0 \ \forall i\in\mathcal{V}$, we have $\sigma > 0$ and the inequality

$$\dot{U} \leq -k_e(\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}) - \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_\iota}})) \leq -k_e(\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}) - \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}})).$$

holds. Thus, if $\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}) - \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}}) > c$, then $\dot{U} \leq -ck_e$ is true. Namely, there exists a finite $\tau(c)$ such that $e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}$ satisfies $\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_\iota}}) - \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}}) < c \ \forall t \geq \tau(c)$ and, from the definition of ι, we also have $\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) - \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}}) < c \ \forall t \geq \tau(c)$ for all $i\in\mathcal{V}$. This completes the proof.

APPENDIX B

PROOF OF LEMMA 2

In the proof, we use the energy functions

$$U_p := \frac{1}{2} \sum_{i \in \mathcal{V}} \|p_i^* - \bar{p}_{io_i}\|^2 = \frac{1}{2} \sum_{i \in \mathcal{V}} \|p^* - \bar{p}_{wo_i}\|^2, \ U_R := \sum_{i \in \mathcal{V}} \phi(e^{-\hat{\xi}\theta_i^*} e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) = \sum_{i \in \mathcal{V}} \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}).$$

We first consider evolution of the position estimates $(\bar{p}_{io_i})_{i \in \mathcal{V}}$ and then show its counterpart with respect to orientation estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}}$ separately. The time evolution of the position estimate $\bar{p}_{io_i}$ in (9) with $V_{wi}^b = 0$ and (16) is described by $\dot{\bar{p}}_{io_i} = k_e(p_{io_i} - \bar{p}_{io_i}) + k_s \sum_{j \in \mathcal{N}_i} (\bar{p}_{io_j} - \bar{p}_{io_i})$. Since the cameras are static, the evolution of $\bar{p}_{wo_i}$ relative to the world frame Σ_w is given by

$$\dot{\bar{p}}_{wo_i} = e^{\hat{\xi}\theta_{wi}} \dot{\bar{p}}_{io_i} = k_e(p_{wo_i} - \bar{p}_{wo_i}) + k_s \sum_{j \in \mathcal{N}_i} (\bar{p}_{wo_j} - \bar{p}_{wo_i}), \tag{32}$$

which is independent of evolution of the orientation estimates (29).

If we define $\bar{q}_i := \bar{p}_{wo_i} - p^*$ and $q_i := p_{wo_i} - p^*$, the time derivative of U_p along with the trajectories of (32) is given by

$$\dot{U}_p = \sum_{i \in \mathcal{V}} \left(k_e \bar{q}_i^T (q_i - \bar{q}_i) + k_s \sum_{j \in \mathcal{N}_i} \bar{q}_i^T (\bar{q}_j - \bar{q}_i) \right)$$

$$= \frac{1}{2} \sum_{i \in \mathcal{V}} \left(k_e (\|q_i\|^2 - \|\bar{q}_i\|^2 - \|q_i - \bar{q}_i\|^2) + k_s \sum_{j \in \mathcal{N}_i} (\|\bar{q}_j\|^2 - \|\bar{q}_i\|^2 - \|\bar{q}_j - \bar{q}_i\|^2) \right).$$

Since $\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \|\bar{q}_j\|^2 - \|\bar{q}_i\|^2 = 0$ holds under Assumption 1 [15], we obtain

$$\dot{U}_p = \frac{1}{2} \sum_{i \in \mathcal{V}} \left(k_e (\|q_i\|^2 - \|\bar{q}_i\|^2 - \|q_i - \bar{q}_i\|^2) - k_s \sum_{j \in \mathcal{N}_i} \|\bar{q}_j - \bar{q}_i\|^2 \right). \tag{33}$$

We see from (33) that if $(\bar{p}_{io_i})_{i \in \mathcal{V}} \in \Omega_p(1)$ then

$$\dot{U}_p \leq -\frac{1}{2} \sum_{i \in \mathcal{V}} \left(k_e \|q_i - \bar{q}_i\|^2 + k_s \sum_{j \in \mathcal{N}_i} \|\bar{q}_j - \bar{q}_i\|^2 \right) \tag{34}$$

holds. From Assumption 2, $\sum_{i \in \mathcal{V}} \|q_i - \bar{q}_i\|^2$ and $\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \|\bar{q}_j - \bar{q}_i\|^2$ are never equal to 0 simultaneously and hence the right-hand side of (34) is strictly negative. Thus, the trajectories of the position estimates $(\bar{p}_{io_i})_{i \in \mathcal{V}}$ along with (32) settle into the set $\Omega_p(1)$ in finite time.

The time derivative of U_R along the trajectories of (29) is given by

$$\dot{U}_R = 2 \sum_{i \in \mathcal{V}} e_R^T (e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) \omega_{uei}, = - \sum_{i \in \mathcal{V}} \text{tr} \left(\text{sk}(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) \hat{\omega}_{uei} \right). \tag{35}$$

Substituting (29) into (35) yields

$$\dot{U}_R = -\sum_{i \in \mathcal{V}} \mathrm{tr}(k_e \Phi_1 + k_s \Phi_2), \tag{36}$$

$$\Phi_1 := \frac{1}{2}(e^{-\hat{\xi}\theta^*} e^{\hat{\xi}\theta_{wo_i}} - e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{-\hat{\xi}\theta_{wo_i}} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}),$$

$$\Phi_2 := \frac{1}{2} \sum_{j \in \mathcal{N}_i} \left(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}} - e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_j}} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} \right).$$

We first consider the term $\sum_i \mathrm{tr}(\Phi_2)$ in (36). From Lemma 6, the following inequality holds.

$$\sum_{i \in \mathcal{V}} \mathrm{tr}(\Phi_2) \geq \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \left\{ \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) - \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) + \sigma_i \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \right\}, \tag{37}$$

where $\sigma_i := \lambda_{min}(\mathrm{sym}(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}))$. Assumption 1 implies that $\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) - \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) = 0$ [15] and hence (37) is rewritten as

$$\sum_{i \in \mathcal{V}} \mathrm{tr}(\Phi_2) \geq \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \sigma_i \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}). \tag{38}$$

We next consider the term $k_e \sum_{i \in \mathcal{V}} \mathrm{tr}(\Phi_1)$ in (36). Applying Lemma 6 again to the term yields

$$\sum_{i \in \mathcal{V}} \mathrm{tr}(\Phi_1) \geq \sum_{i \in \mathcal{V}} \left\{ \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) - \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\xi}\theta_{wo_i}}) + \sigma_i \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\xi}\theta_{wo_i}}) \right\}. \tag{39}$$

Substituting (38) and (39) into (36) yields

$$\dot{U}_R \leq -\sum_{i \in \mathcal{V}} \left(k_e \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) - k_e \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\xi}\theta_{wo_i}}) \right.$$
$$\left. + \sigma_i \left(k_e \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\xi}\theta_{wo_i}}) + k_s \sum_{j \in \mathcal{N}_i} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \right) \right). \tag{40}$$

If $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}} \notin \Omega_R(1)$ is true, (40) is rewritten as

$$\dot{U}_R \leq -\sum_{i \in \mathcal{V}} \sigma_i \left(k_e \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\xi}\theta_{wo_i}}) + k_s \sum_{j \in \mathcal{N}_i} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \right). \tag{41}$$

Note that, from the assumption of $e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} > 0$, we have $\sigma_i > 0$. Since both of the terms $\sum_{i \in \mathcal{V}} \sigma_i \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\xi}\theta_{wo_i}})$ and $\sum_{i \in \mathcal{V}} \sigma_i \sum_{j \in \mathcal{N}_i} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}})$ are never equal to 0 under Assumption 2, the right-hand side of (41) is strictly negative. This implies that the trajectories of the estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}}$ converge to the set $\Omega_R(1)$ in finite time.

APPENDIX C

PROOF OF LEMMA 3

Suppose that $\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) < \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}}) + c$ holds true for some $c > 0$. Then, from Hoff-man-Wielandt's perturbation theorem [33], we have

$$\left|\lambda_{min}(\text{sym}(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})) - \lambda_{min}(\text{sym}(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta^*}))\right| \leq \left\|\text{sym}(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) - \text{sym}(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta^*})\right\|_F$$

$$\leq \|e^{-\hat{\xi}\theta^*}(e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} - e^{\hat{\xi}\theta^*})\|_F = \|e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} - e^{\hat{\xi}\theta^*}\|_F < \sqrt{2(\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}}) + c)}.$$

This immediately means

$$\lambda_{min}(\text{sym}(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})) \geq \beta := 1 - \sqrt{2(\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}}) + c)}. \tag{42}$$

From Lemma 1 and (42), Inequality (40) is rewritten as

$$\dot{U}_R \leq -\sum_{i\in\mathcal{V}}\left(k_e\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) - k_e\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_i}})\right.$$

$$\left. + \beta\left(k_e\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\xi}\theta_{wo_i}}) + k_s\sum_{j\in\mathcal{N}_i}\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}})\right)\right) \tag{43}$$

at least after the time $\tau(c)$. If $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \in \mathcal{S}_2^R(k)$ holds true, then we have

$$\dot{U}_R \leq -k_e\sum_{i\in\mathcal{V}}\left(\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) + \beta\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\xi}\theta_{wo_i}})\right) \tag{44}$$

at least after the time $\tau(c)$. Under Assumption 2, the right-hand side of (44) is strictly negative.

In terms of U_p, from (33), if $(\bar{p}_{io_i})_{i\in\mathcal{V}} \in \mathcal{S}_2^p(k)$, we have $\dot{U}_p = -\frac{k_e}{2}\sum_{i\in\mathcal{V}}\left(\|\bar{q}_i\|^2 + \|q_i - \bar{q}_i\|^2\right)$, whose right-hand side is strictly negative under Assumption 2. These complete the proof.

APPENDIX D

PROOF OF THEOREM 1

We first consider evolution of the position estimates $(\bar{p}_{io_i})_{i\in\mathcal{V}}$ described by (32). The case not satisfying $k \leq 1/W$ is already proved in Lemma 2 and hence we consider the case such that $k \leq 1/W$ is satisfied. Lemmas 2 and 3 indicate that $\dot{U}_p < 0$ holds in $\Omega_p(1) \cup \mathcal{S}_2^p(k)$. Namely, from the inclusion (21), we have $\dot{U}_p < 0$ except for the region $\Omega_p(\varepsilon_p)$ if $\dot{U}_p < 0$ holds in the region $\mathcal{S}_3^p(k, \varepsilon_p)$. If it is true, the trajectories along with (32) settle into the set $\Omega_p(\varepsilon_p)$ in finite time. It is thus sufficient to prove that $\dot{U}_p$ is strictly negative for all $(\bar{p}_{io_i})_{i\in\mathcal{V}} \in \mathcal{S}_3^p(k, \varepsilon_p)$.

Equation (33) is rewritten as

$$\dot{U}_p = -\frac{k_e}{2} \sum_{i \in \mathcal{V}} \Big(-\|q_i\|^2 + \|\bar{q}_i\|^2 + (1-\epsilon)\|q_i - \bar{q}_i\|^2 \Big) - a_p, \tag{45}$$

where $a_p := \frac{1}{2} \sum_{i \in \mathcal{V}} \Big(k_e \epsilon \|q_i - \bar{q}_i\|^2 + k_s \sum_{j \in \mathcal{N}_i} \|\bar{q}_j - \bar{q}_i\|^2 \Big)$ is strictly positive under Assumption 2. Now, for any $\alpha \in (0,1)$ and $j^* \in \mathcal{V}$, we have

$$\|q_i - \bar{q}_i\|^2 \geq \alpha\|q_i - \bar{q}_{j^*}\|^2 - \frac{\alpha}{1-\alpha}\|\bar{q}_i - \bar{q}_{j^*}\|^2. \tag{46}$$

Let j^* be a node satisfying $j^* = \arg\min_{i_0} D(i_0)$ and $G_T^* = (\mathcal{V}, \mathcal{E}_T^*) \in \mathcal{T}(j^*)$ be a graph satisfying $G_T^* = \arg\min_{G_T \in \mathcal{T}(j^*)} \tilde{D}(G_T)$, where D and $\tilde{D}$ are defined in (4). Then, we obtain

$$\|\bar{q}_i - \bar{q}_{j^*}\|^2 = \Big\| \sum_{l \in \{0, \cdots, d_{G_T^*}(i)-1\}} (\bar{q}_{v_l(i)} - \bar{q}_{v_{l+1}(i)}) \Big\|^2 \leq d_{G_T^*}(i) \sum_{l \in \{0, \cdots, d_{G_T^*}(i)-1\}} \|\bar{q}_{v_l(i)} - \bar{q}_{v_{l+1}(i)}\|^2,$$

where $(v_0(i), \cdots, v_{d_{G_T^*}(i)-1}(i))$ is the path from root j^* to node i along tree G_T^*. Namely,

$$\sum_{i \in \mathcal{V}} \|\bar{q}_i - \bar{q}_{j^*}\|^2 \leq \sum_{i \in \mathcal{V}} d_{G_T^*}(i) \sum_{l \in \{0, \cdots, d_{G_T^*}(i)-1\}} \|\bar{q}_{v_l(i)} - \bar{q}_{v_{l+1}(i)}\|^2. \tag{47}$$

holds. For any edge $E = (v^1, v^2)$ of G_T^*, the coefficient of $\|\bar{q}_{v^1} - \bar{q}_{v^2}\|^2$ in the right hand side of (47) is given by $\sum_{i \in \mathcal{V}} \delta_{G_T^*}(E; i) d_{G_T^*}(i)$, which is upper-bounded by $\tilde{D}(G_T^*) = W$. We thus have

$$\sum_{i \in \mathcal{V}} \|\bar{q}_i - \bar{q}_{j^*}\|^2 \leq W \sum_{E=(v^1, v^2) \in \mathcal{E}_T^*} \|\bar{q}_{v^1} - \bar{q}_{v^2}\|^2 \leq W \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \|\bar{q}_i - \bar{q}_j\|^2. \tag{48}$$

The latter inequality of (48) holds because G_T^* is a subgraph of G_u. Since $(\bar{p}_{io_i})_{i \in \mathcal{V}} \in \mathcal{S}_3^p(k, \varepsilon_p)$, the inclusion $(\bar{p}_{io_i})_{i \in \mathcal{V}} \notin \mathcal{S}_2^p(k)$ holds and hence

$$\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \|\bar{q}_i - \bar{q}_j\|^2 = \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \|\bar{p}_{wo_i} - \bar{p}_{wo_j}\|^2 \leq 2k\rho_p. \tag{49}$$

Moreover, the following inequality holds from the definition of the average p^*.

$$\sum_{i \in \mathcal{V}} \|q_i - \bar{q}_{j^*}\|^2 = \sum_{i \in \mathcal{V}} \|p_{wo_i} - \bar{p}_{wo_{j^*}}\|^2 \geq \sum_{i \in \mathcal{V}} \|p_{wo_i} - p^*\|^2 = 2\rho_p \tag{50}$$

From (46), (48), (49) and (50), equation (45) is rewritten as

$$\dot{U}_p \leq k_e \left\{ -\frac{1}{2} \Big(\sum_{i \in \mathcal{V}} \|\bar{q}_i\|^2 \Big) + \Big(1 - (1-\epsilon)\Big(\alpha - \frac{kW\alpha}{1-\alpha} \Big) \Big) \rho_p \right\} - a_p. \tag{51}$$

If $(\bar{p}_{io_i})_{i \in \mathcal{V}} \in \mathcal{S}_3^p(k, \varepsilon_p)$, then $(\bar{p}_{io_i})_{i \in \mathcal{V}} \notin \Omega_p(\varepsilon_p)$ and hence (51) is rewritten as

$$\dot{U}_p \leq k_e \left\{ (1-\epsilon)\Big(\alpha - \frac{kW\alpha}{1-\alpha} - \big(1 - \sqrt{kW}\big)^2 \Big) \right\} \rho_p - a_p.$$

Under the assumption that $k \leq 1/W$, the inequality $\alpha - \frac{kW\alpha}{1-\alpha} \leq \left(1 - \sqrt{kW}\right)^2$ holds for any $\alpha \in (0,1)$ and hence $\dot{U}_p \leq -a_p < 0$. This completes the proof of the former half of the theorem.

We next consider the evolution of the orientation estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}}$ described by (29). The case not satisfying $k \leq \beta/W$ or $\beta > 0$ is already proved in Lemma 2. We thus consider the case such that $k \leq \beta/W$ and $\beta > 0$ hold. We first note that the set $\mathcal{S} = \{(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}}|\ e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0 \ \forall i \in \mathcal{V}\}$ is a positively invariant set from Lemma 1 and hence trajectories of $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}}$ starting from $\mathcal{S}$ never gets out of $\mathcal{S}$. Lemmas 2 and 3 also prove that, in the region $\mathcal{S}$, $\dot{U}_R < 0$ holds if $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}} \in (\mathcal{S} \setminus \Omega_R(1)) \cup \mathcal{S}_2^R(k)$ at least after the time $\tau(c)$. Namely, as long as $\dot{U}_R < 0$ is true in the region $\mathcal{S}_3^R(k, \varepsilon_R)$, the inequality $\dot{U}_R < 0$ holds except for the region $\Omega_R(\varepsilon_R)$ from the inclusion (21), which means the trajectories along with (29) settle into the set $\Omega_R(\varepsilon_R)$ in finite time. It is thus sufficient to prove that $\dot{U}_R$ is strictly negative for all $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}} \in \mathcal{S}_3^R(k, \varepsilon_R)$.

We first notice that if we define $a_R := \beta \sum_{i \in \mathcal{V}} \left(k_e \epsilon \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\xi}\theta_{wo_i}}) + k_s \sum_{j \in \mathcal{N}_i} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \right)$, a_R is strictly positive under Assumption 2. Using the parameter a_R, (43) is rewritten as

$$\dot{U}_R \leq -k_e \sum_{i \in \mathcal{V}} \left(-\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_i}}) + \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) + \beta(1-\epsilon)\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\xi}\theta_{wo_i}}) \right) - a_R. \quad (52)$$

We thus consider the former three terms of the right hand side of Inequality (52). We first have

$$\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\xi}\theta_{wo_i}}) \geq \alpha\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_{j^*}}}e^{\hat{\xi}\theta_{wo_i}}) - \frac{\alpha}{1-\alpha}\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_{j^*}}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) \quad (53)$$

for any $\alpha \in (0,1)$ and $j^* \in \mathcal{V}$. Again, let j^* be a node satisfying $j^* = \arg\min_{i_0} D(i_0)$ and $G_T^* = (\mathcal{V}, \mathcal{E}_T^*) \in \mathcal{T}(j^*)$ be a graph satisfying $G_T^* = \arg\min_{G_T \in \mathcal{T}(j^*)} \tilde{D}(G_T)$. Then, the inequality

$$\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_{j^*}}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) \leq d_{G_T^*}(i) \sum_{l \in \{0, \cdots, d_{G_T^*}(i)-1\}} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_{v_l(i)}}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_{v_{l+1}(i)}}}) \quad (54)$$

holds from the definition of the energy function ϕ and hence

$$\sum_{i \in \mathcal{V}} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_{j^*}}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) \leq \sum_{i \in \mathcal{V}} d_{G_T^*}(i) \sum_{l \in \{0, \cdots, d_{G_T^*}(i)-1\}} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_{v_l(i)}}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_{v_{l+1}(i)}}}). \quad (55)$$

Similarly to the case of position estimates, (55) is rewritten as

$$\sum_{i \in \mathcal{V}} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_{j^*}}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) \leq W \sum_{E=(v^1,v^2) \in \mathcal{E}_T} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_{v^1}}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_{v^2}}}) \leq W \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}). \quad (56)$$

Since $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}} \in \mathcal{S}_3^R(k, \varepsilon_R)$, the inclusion $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i \in \mathcal{V}} \notin \mathcal{S}_2^R(k)$ holds and hence

$$\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \leq \frac{k\rho_R}{\beta} \quad (57)$$

is true. We next focus on $\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_j*}}e^{\hat{\xi}\theta_{wo_i}})$ in (53). From the definition of the average $e^{\hat{\xi}\theta^*}$ (7),

$$\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_j*}}e^{\hat{\xi}\theta_{wo_i}})\geq\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_i}})=\rho_R \tag{58}$$

holds for any $e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}\in SO(3)$. Substituting (53), (56), (57) and (58) into inequality (52) yields

$$\dot{U}_R\leq-k_e\Big\{\Big(\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})\Big)-\Big(1-(1-\epsilon)\Big(\alpha\beta-\frac{kW\alpha}{1-\alpha}\Big)\Big)\rho_R\Big\}-a_R. \tag{59}$$

If $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}\in\mathcal{S}_3^R(k,\varepsilon_R)$, then $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}\notin\Omega(\varepsilon_R)$ and hence (59) is rewritten by

$$\dot{U}_R\leq k_e(1-\epsilon)\Big(\alpha\beta-\frac{kW\alpha}{1-\alpha}-\Big(\sqrt{\beta}-\sqrt{kW}\Big)^2\Big)\rho_R-a_R. \tag{60}$$

Let us now notice that, under $k\leq\beta/W$, $\alpha\beta-\frac{kW\alpha}{1-\alpha}\leq\Big(\sqrt{\beta}-\sqrt{kW}\Big)^2$ holds true for any $\alpha\in(0,1)$ and hence $\dot{U}_R\leq-a_R<0$. This completes the proof of the latter half of the theorem.

APPENDIX E

PROOF OF LEMMA 5

Suppose that $S(t)$ moves from $S(t)=S$ to $S(t+t_\Delta)=S+\Delta S$. We also describe $e^{\hat{\xi}\theta^*}(t+t_\Delta)$ as $e^{\hat{\xi}\theta^*}(t+t_\Delta)=e^{\hat{\xi}\theta^*}(t)+\Delta e^{\hat{\xi}\theta^*}$, $\Delta e^{\hat{\xi}\theta^*}:=\text{Proj}(S+\Delta S)-e^{\hat{\xi}\theta^*}(t)$. Then, if $\|\Delta S\|_F^2\leq\bar{s}$ is true for some $\bar{s}$, it is proved in [34] that under (26)

$$\sup_{\Delta S}\|\Delta e^{\hat{\xi}\theta^*}\|_F^2\leq b:=4(1-(1-\mu^2(\gamma)\bar{s}/2)^{1/2})<\mu^2(\gamma)\bar{s}. \tag{61}$$

The hypothesis of $\|\Delta e^{\hat{\xi}\theta^*}\|_F^2\geq\mu^2(\gamma)\|\Delta S\|_F^2$ contradicts (61) and hence $\|\Delta e^{\hat{\xi}\theta^*}\|_F^2<\mu^2(\gamma)\|\Delta S\|_F^2$. From continuous differentiability of the average $e^{\hat{\xi}\theta^*}$, we also get

$$\|\omega^{b,*}\|^2=\|\dot{e}^{\hat{\xi}\theta^*}\|_F^2=\|\lim_{t_\Delta\to0}\Delta e^{\hat{\xi}\theta^*}/t_\Delta\|_F^2=\lim_{t_\Delta\to0}\|\Delta e^{\hat{\xi}\theta^*}/t_\Delta\|_F^2$$
$$<\lim_{t_\Delta\to0}\mu^2(\gamma)\|\Delta S/t_\Delta\|_F^2=\mu^2(\gamma)\|\lim_{t_\Delta\to0}\Delta S/t_\Delta\|_F^2=\mu^2(\gamma)\|\dot{S}\|_F^2.$$

It is clear that $n\|\dot{S}(t)\|_F^2\leq\|w(t)\|^2$ holds and hence (27) is true.

APPENDIX F

PROOF OF THEOREM 2

We first consider the statement in terms of the position estimates. The time derivative of U_p along the trajectories of the system Σ_{track} is given by

$$\dot{U}_p=\sum_{i\in\mathcal{V}}\bar{q}_i^T(e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}v_{ue}-e^{\hat{\xi}\theta^*}v^{b,*}). \tag{62}$$

From Lemma 2, we obtain

$$\sum_{i\in\mathcal{V}}\bar{q}_i^T e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}v_{ue} < \frac{k_e}{2}\sum_{i\in\mathcal{V}}(\|q_i\|^2 - \|\bar{q}_i\|^2) \le k_e\rho_p' - \frac{k_e}{2}\sum_{i\in\mathcal{V}}\|\bar{q}_i\|^2 \tag{63}$$

under Assumptions 1 and 3. In addition, under Assumption 3, the second term of (62) satisfies

$$-\sum_{i\in\mathcal{V}}\bar{q}_i^T e^{\hat{\xi}\theta^*}v^{b,*} = \frac{1}{2}\sum_{i\in\mathcal{V}}\Big(-\|\bar{q}_i - e^{\hat{\xi}\theta^*}v^{b,*}\|^2 + \|\bar{q}1\|^2 + \|e^{\hat{\xi}\theta^*}v^{b,*}\|^2\Big) \le \frac{1}{2}\sum_{i\in\mathcal{V}}\Big(\|\bar{q}_i\|^2 + \|v^{b,*}\|^2\Big)$$

$$\le \frac{1}{2}\Big(\sum_{i\in\mathcal{V}}\|\bar{q}_i\|^2\Big) + \frac{n}{2}\|v^{b,*}\|^2 \le \frac{1}{2}\Big(\sum_{i\in\mathcal{V}}\|\bar{q}_i\|^2\Big) + \frac{1}{2}\|w_p\|^2 \le \frac{1}{2}\Big(\sum_{i\in\mathcal{V}}\|\bar{q}_i\|^2\Big) + \frac{1}{2}\bar{w}_p^2. \tag{64}$$

Substituting (63) and (64) into (62) yields

$$\dot{U}_p < \sum_{i\in\mathcal{V}}k_e\rho_p' - (k_e - 1)\Big(\frac{1}{2}\sum_{i\in\mathcal{V}}\|\bar{q}_i\|^2\Big) + \frac{1}{2}\bar{w}_p^2. \tag{65}$$

Now, we see from (65) and the definition of ε_p' that $\dot{U}_p < 0$ as long as $(\bar{p}_{io_i})_{i\in\mathcal{V}} \notin \Omega_p'(\varepsilon_p')$. Hence, the function $U_p(t)$ is monotonically strictly decreasing in the region and there exists a finite time T such that $(\bar{p}_{io_i})_{i\in\mathcal{V}} \in \Omega_p'(\varepsilon_p')$ $\forall t \ge T$.

We next consider the evolution of orientation estimates. The time derivative of U_R along the trajectories of the system Σ_{track} is given by

$$\dot{U}_R = 2\sum_{i\in\mathcal{V}}e_R^T(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})(\omega_{uei} - \omega^{b,*}). \tag{66}$$

From Lemma 2, we obtain

$$2\sum_{i\in\mathcal{V}}e_R^T(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})\omega_{uei} < -k_e\sum_{i\in\mathcal{V}}\Big(\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) - \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_i}})\Big)$$

$$\le k_e\rho_R' - k_e\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) \tag{67}$$

under the assumption of $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(t))_{i\in\mathcal{V}} \in \mathcal{S}$ $\forall t \ge 0$ and Assumptions 1 and 3. We also have

$$-2\sum_{i\in\mathcal{V}}e_R^T(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})\omega^{b,*}$$

$$= -\sum_{i\in\mathcal{V}}\Big(\frac{1}{\mu^2(\gamma)}\|\mu^2(\gamma)e_R(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) + \omega^{b,*}\|^2 - \mu^2(\gamma)\|e_R(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})\|^2\Big) + \frac{n}{\mu^2(\gamma)}\|\omega^{b,*}\|^2$$

$$\le \sum_{i\in\mathcal{V}}\Big(\mu^2(\gamma)\|e_R(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})\|^2\Big) + \frac{n}{\mu^2(\gamma)}\|\omega^{b,*}\|^2 \le \sum_{i\in\mathcal{V}}\Big(\mu^2(\gamma)\|e_R(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})\|^2\Big) + \bar{w}_R^2, \tag{68}$$

where the last inequality holds from Lemma 5. Since $\|e_R(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})\|^2 \leq \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})$ is true, substituting (67) and (68) into (66) yields

$$\dot{U}_R < k_e\rho'_R + \bar{w}_R^2 - (k_e - \mu^2(\gamma)) \sum_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}).$$

Now, if $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \notin \Omega'_R(\varepsilon'_R)$ holds, then $\dot{U}_R < 0$. Hence, the function $U_R(t)$ is monotonically strictly decreasing in the region and this completes the proof.

REFERENCES

[1] H. Aghajan and A. Cavallaro (Eds), "Multi-Camera Networks: Principles and Applications," Academic Press, 2009.

[2] M. Zhu and S. Martinez, "Distributed Coverage Games for Mobile Visual Sensors (I), Reaching the set of Nash equilibria," Proc. of the 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, pp. 169–174, 2009.

[3] Y. Ma, S. Soatto, J. Kosecka and S. S. Sastry, "An Invitation to 3-D Vision: From Images to Geometric Models," Springer, 2004.

[4] P. A. Vela and I. J. Ndiour, "Estimation Theory and Tracking of Deformable Objects," Proc. of the 2010 IEEE Multi-conference on Systems and Control, pp. 1222–1233, 2010.

[5] M. Sznaier and O. Camps, "Dynamics Based Extraction of Information Sparsely Encoded in High Dimensional Data Streams," Proc. of the 2010 IEEE Multi-conference on Systems and Control, pp. 1234–1245, 2010.

[6] N. R. Gans and S. A. Hutchinson, "Stable Visual Servoing through Hybrid Switched System Control," IEEE Trans. on Robotics, Vol. 23, No. 3, pp. 530–540, 2007.

[7] A. P. Dani, N. R. Fischer, K. Zhen and W. E. Dixon, "Nonlinear Observer for Structure Estimation Using A Paracatadioptric Camera," Proc. of the 2010 American Control Conference, pp. 3487–3492, 2010.

[8] M. Fujita, H. Kawai and M. W. Spong, "Passivity-based Dynamic Visual Feedback Control for Three Dimensional Target Tracking:Stability and L2-gain Performance Analysis," IEEE Trans. on Control Systems Technology, Vol. 15, No. 1, pp. 40–52, 2007.

[9] T. Hatanaka and M. Fujita, "Passivity-based Visual Motion Observer: From Theory to Distributed Algorithms," Proc. of the 2010 IEEE Multi-conference on Systems and Control, pp. 1210–1221 , 2010.

[10] R. Olfati-Saber, J. A. Fax and R. M. Murray, "Consensus and Cooperation in Networked Multi-Agent Systems," Proc. of the IEEE, Vol. 95, No. 1, pp. 215–233, 2007.

[11] F. Bullo, J. Cortes and S. Martinez, "Distributed Control of Robotic Networks," Princeton University Press, 2009.

[12] M. Arcak, "Passivity as a Design Tool for Group Coordination," IEEE Trans. on Automatic Control, Vol. 52, No. 8, pp. 1380–1390, 2007.

[13] N. Chopra and M. W. Spong, "Passivity-Based Control of Multi-Agent Systems," Advances in Robot Control: From Everyday Physics to Human-Like Movements, S. Kawamura and M. Svnin (eds.), pp. 107–134, Springer, 2006.

[14] H. Yu, F. Zhu and P. J. Antsaklis, "Event-Triggered Cooperative Control for Multi-Agent Systems Based on Passivity Analysis," ISIS Technical Report, 2010.

[15] Y. Igarashi, T. Hatanaka, M. Fujita and M. W. Spong, "Passivity-based Attitude Synchronization in $SE(3)$," IEEE Trans. on Control Systems Technology, Vol. 17, No. 5, pp. 1119–1134, 2009.

[16] L. Xiao, S. Boyd and S. Lall, "A Scheme for Robust Distributed Sensor Fusion Based on Average Consensus," Proc. of the International Conference on Information Processing in Sensor Networks, pp. 63–70, 2005.

[17] R. Tron, R. Vidal and A. Terzis, "Distributed Pose Averaging in Camera Sensor Networks via Consensus on SE(3)," Proc. of the International Conference on Distributed Smart Cameras, 2008.

[18] R. Olfati-Saber, "Distributed Kalman Filter with Embedded Consensus Filters," Proc. of the 44th IEEE Conference on Decision and Control and 2005 European Control Conference, pp. 8179–8184, 2005.

[19] R. Olfati-Saber and J. S. Shamma, "Consensus Filters for Sensor Networks and Distributed Sensor Fusion," Proc. of the 44th IEEE Conference on Decision and Control and 2005 European Control Conference, pp. 6698–6703, 2005.

[20] R. Olfati-Saber, "Distributed Kalman Filter for Sensor Networks," Proc. of the 46th IEEE Conference on Decision and Control, pp. 5492–5498, 2007.

[21] R. Carli, A. Chiuso, L. Schenato and S. Zampieri, "Distributed Kalman Filtering Using Consensus Strategies," Proc. of the 46th IEEE Conference on Decision and Control, pp. 5486–5491, 2007.

[22] U. A. Khan and J. M. F. Moura, "Distributing The Kalman filter for Large-scale Systems," IEEE Trans. on Signal Processing, Vol. 56, No. 10, pp. 4919–4935, 2008.

[23] H. Bai and R. A. Freeman and K. M. Lynch "Robust Dynamic Average Consensus of Time-varying Inputs," Proc. of the 49th IEEE Conference on Decision and Control, pp. 3104–3109, 2010.

[24] U. A. Khan, S. Kar, A. Jadbabaie, J. M. F. Moura, "On Connectivity, Observability, and Stability in Distributed Estimation," Proc. of the 49th IEEE Conference on Decision and Control, pp. 3104–3109, 2010.

[25] M. Moakher, "Means and Averaging in the Group of Rotations," SIAM Journal on Matrix Analysis and Applications, Vol. 24, No. 1, pp. 1–16, 2002.

[26] J. H. Manton, "A Globally Convergent Numerical Algorithm for Computing the Centre of Mass on Compact Lie Groups," Proc. of the 8th Control, Automation, Robotics and Vision Conference, Vol. 3, pp. 2211–2216, 2004.

[27] T. Hatanaka and M. Fujita "Passivity-based Cooperative Estimation of 3D Target Motion for Visual Sensor Networks: Analysis on Averaging Performance," Proc. of 2011 American Control Conference, to appear, 2011.

[28] T. Hatanaka, K. Hirata and M. Fujita, "Cooperative Estimation of 3D Target Object Motion via Networked Visual Motion Observers," Proc. of The 50th IEEE Conference on Decision and Control and European Control Conference, submitted, 2011. (available at http://www.fl.ctrl.titech.ac.jp/paper/2011/HHF_CDC11.pdf)

[29] H. Kawai, T. Murao and M. Fujita, "Passivity-based Visual Motion Observer with Panoramic Camera for Pose Control," Journal of Intelligent and Robotic Systems, to appear, 2011. (available at http://www.springerlink.com/content/k5m8u11668332800/fulltext.pdf)

[30] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, "SURF: Speeded Up Robust Features," Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346–359, 2008.

[31] P. A. Absil, R. Mahony and R. Sepulchre, "Optimization Algorithms on Matrix Manifolds," Princeton University Press, 2008.

[32] A. Nedic and A. Ozdaglar, "Distributed Subgradient Methods for Multi-agent Optimization," IEEE Trans. on Automatic Control, Vol. 54, No. 1, pp. 48-61, 2009.

[33] L. Elsner and S. Friedland, "Singular Values, Doubly Stochastic Matrices, and Applications," Linear Algebra and Its Applications, Vol. 220, pp. 161–169, 1995.

[34] I. Soderkvist, "Perturbation Analysis of the Orthogonal Procrustes Problem," BIT Numerical Mathematics, Springer, Vol. 33, No. 4, pp. 687–694, 1993.